Understanding Quantum Phenomena

Soazig Le Bihan

Université Nancy 2 - Henri Poincaré – Université de Bielefeld

LPHS – Archives Poincaré – CNRS UMR 7117

Ecole Doctorale Langage, Temps, Société

Comprendre les phénomènes quantiques

Thèse

Pour obtenir le diplôme de:
Docteur de Philosophie

Présentée et soutenue par:
Soazig Le Bihan

le 13 Février 2008

Sous la direction de:
Martin Carrier et Gerhard Heinzmann

Président du Jury:	Prof. Guido Bacciagaluppi *Université de Sydney*	
Membres du Jury:	Prof. Harvey Brown *Université d'Oxford*	DR. Michel Bitbol *CREA / Ecole Polytechnique, Université Paris 1 Sorbonne*
Prof. Martin Carrier *Université de Bielefeld*	Prof. Michael Detlefsen *Université Notre Dame*	Prof. Gerhard Heinzmann *Université Nancy 2*

University of Nancy 2 - Henri Poincaré – University of Bielefeld

LPHS – Archives Poincaré – CNRS UMR 7117

Understanding Quantum Phenomena

Ph.D. Thesis

Submitted to the jury in fulfillment of the requirements for the degree of:
Doctor of Philosophy

Presented by:
Soazig Le Bihan

Defended on: February 13th, 2008

Under the supervision of:
Martin Carrier and Gerhard Heinzmann

Contents

Acknowledgments

I wish to thank Pierre Bourdon, Francis Wolff, Anouk Barberousse, Claude Imbert, Pierre Zaoui, Daniel Andler and Bas van Fraassen who all convinced me at various turning points of my life that aiming at the highest goals was really worth trying despite what I believed I was capable of achieving. Along the same lines, I am also greatly indebted to Feisal Ben Hassel for his inalterable trust and support.

I wish to thank Gerhard Heinzman for his infatigable but always lighthearted support during the completion of this work.

I wish to express my deepest gratitude to Guido Bacciagaluppi for continuous intellectual feedback and moral support.

Finally, I owe too much to Armond Duwell to even be able to express it. I dedicate this work to the new possibilities he opened for our future.

This work was made possible thanks to the financial support of many institutions. I was able to pursue studies at the university thanks to the Ecole Normale Supérieure. I was able to spend a year at the Philosophy Department in Princeton University thanks to a Jane Eliza Procter Fellowship. The Ministère de la Recherche et de l'Enseignement Supérieur allowed me an Allocation Couplée, together with the Université Nancy 2 and the Université Paris VII – Jussieu, for working on my dissertation. Thanks to Prof. Gerhard Heinzman and Prof.

Martin Carrier, who together put a cotutelle in place for me, I was able to apply and get a Bourse de Mobilité from the European Union, which allowed me to attend many international conferences. I am also grateful to the Association des Femmes Françaises Diplômées de l'Université for allowing me a Bourse Huguette Delavault in 2005. I received financial support from the University of Bielefeld, the DAAD, and the Université Franco-Allemande, thanks to the help of Prof. Martin Carrier. Finally, I could finish my dissertation thanks to the Sawyier Predoctoral Fellowship at the Illinois Institute of Technology of Chicago. Many thanks to all of the people who helped me through all this.

Three characterizations of philosophy of physics

At a recent public lecture, Harvard physicist and University Professor Emeritus Edward Purcell told his audience of his happiness that he has lived to see a philosophical problem settled in the laboratory. [...] Professor Purcell's remark that day was prompted by his assessment of the astonishing degree to which recent experiments bear directly on issues of a traditionally philosophical character. I, too, am happy that this situation to which Purcell referred has taken place in my lifetime. (J. Jarrett [73, p. 60])

It is legitimate, in view of all these rich results, to speak of the enterprise of experimental metaphysics*. A warning is needed, however, against possible misunderstanding of this term. One should not anticipate straightforward and decisive resolution of metaphysical disputes by the outcomes of experiments. We know, in fact, that laboratory tests alone do not settle without careful conceptual analysis even those problems which are commonly classified as scientific, and* a fortiori *such analysis is indispensable in coming to grips with metaphysical problems.* (A. Shimony, [103, p. 29])

The moral is, it seems to me, that scientists who constructed quantum mechanics did exactly as much as we – or science – needs. It is possible to gain greater insight into the structure of that theory [...]. It is not possible to reach independent insight into the "fundamental" structure of nature and to illuminate quantum theory by viewing it sub specie *that structure. But then, that is not needed. There does not have to be a reason for everything.* (B. van Fraassen, [130, p. 113])

Résumé

Les théories physiques classiques peuvent être interprétées comme représentant le monde comme constitué de systèmes characterisés par des propriétés bien définies, et interagissant de façon causale, locale, et déterministe. Il est bien connu que la mécanique quantique orthodoxe, qui se trouve être par ailleurs une de nos théories les mieux confirmées, ne peut pas être interprétée dans les termes du cadre classique décrit ci-dessus. Les théorèmes de type Bell, and les expériences qui y sont associées, ont aggravé la situation en rendant une telle interprétation impossible. Au début des années soixante, John Bell a démontré que toute théorie représentant son domaine dans les termes du cadre classique ci-dessus satisfait une série d'inégalités, les inégalités de Bell. L'expérience montre que les phénomènes quantiques violent ces inégalités. Par un simple modus tollens, il suit de ceci qu'aucune théorie qui représente le monde en termes classiques ne peut être compatible avec tous les phénomènes quantiques.

Certains physiciens ont développé des théories alternatives à la théorie quantique orthodoxe, théories dont les interprétations associées abandonnent au moins une des caractéristiques du cadre classique. Les philosophes s'efforcent d'interpréter ce résultat, i.e. s'efforcent de comprendre à quoi peut ressembler le monde s'il est vrai que les interactions causales entre les systèmes physiques sont soit non-locales, soit non-déterministes, ou bien que ces systèmes physiques ne possèdent pas de propriétés définies. Cette ligne de pensée a atteint son apogée avec la "métaphysique expérimentale" qui s'est développée après

que la violation des inégalités de Bell par l'expérience a été observée. Les défenseurs de la métaphysique expérimentale soutiennent que l'on peut tirer des conclusions métaphysiques de nos meilleures théories et des expériences qui leur sont associées. L'interprétation orthodoxe est que les expériences de type Bell nous forcent à accepter certaines formes d'interactions non-locales, et peut-être non causales, au niveau fondamental.

Dans cette thèse, j'évalue dans quelle mesure l'investigation philosophique de nos meilleures théories scientifiques peut nous aider à décider de quoi est fait le monde au niveau fondamental. J'étudie cette question en considérant le domaine quantique, et en particulier les phénomènes de type Bell. Mes conclusions vont dans le sens d'une vue plus modeste du rôle possible de la philosophie de la physique que celle que les programmes de métaphysique expérimentale suggèrent.

Dans une première partie, j'étudie le rôle que la philosophie de la physique peut légitimement jouer dans le développement et l'évaluation des différentes theories quantiques. Je soutiens que le rôle de la philosophie de la physique ne consiste pas à imposer des critères d'acceptabilité pour les théories physiques en plus des critères usuels de cohérence interne et d'adéquation empirique. Un rôle à la fois important et légitime que peut jouer la philosophie de la physique consiste à distinguer clairement, pour nos théories et leurs interprétations, d'un côté ce qui est imposé par les phénomènes et les théories, et de l'autre ce qui relèvent de nos choix et préférences. Je soutiens enfin qu'un tel travail de distinction peut se faire sur la base de l'analyse structurelle de nos théories et de nos modèles de données.

Dans la seconde partie de la thèse, j'étudie spécifiquement le cas de l'interprétation des théorèmes de Bell et des expériences associées. J'évalue systématiquement l'interprétation orthodoxe selon laquelle les résultats de ces expériences nous forcent à accepter une forme "bénigne" de non-localité, "bénigne" parce que de type non-causal. Je montre que l'interprétation orthodoxe comprend trois thèses distinctes, la première portant sur la question de la localité, la seconde portant sur la question de la causalité et la troisième concernant le holisme, cette dernière dependant directement de la deuxième. Je fournis des cadres théoriques rigoureux afin d'évaluer les deux premières thèses. Le résultat de mon analyse concernant la question de la localité est que l'interprétation orthodoxe peut être défendue. En revanche, concernant les questions de causalité, je montre qu'aucun des cadres théoriques existant ne permet de soutenir l'interprétation orthodoxe comprise dans sa version métaphysique forte. Puisque la thèse concernant la causalité ne peut être soutenue, la dernière thèse, concernant le holisme, ne le peut non plus. Dans les deux cas cependant, une version faible peut cependant être défendue. En particulier, un résultat important de ma thèse est que l'interprétation orthodoxe peut être défendue rigoureusement au niveau empirique.

Abstract

It so happens that classical physical theories can be interpreted as a representation of local, deterministic, causal interactions between systems with definite properties. Orthodox quantum theory, which happens to be one of our most experimentally well-confirmed theories, is notoriously resistant to being interpreted in terms of the above framework. Bell-type theorems and Bell-type experiments have made such an interpretation impossible. In the early sixties, John Bell demonstrated that any theory that represents its domain in terms of the above framework satisfies a set of inequalities, the so-called Bell inequalities. Experiments on quantum phenomena violate Bell-type inequalities. By a simple modus tollens, the upshot is that no theory that includes all the elements of the above framework can give an account of all quantum phenomena.

Philosophers have been trying to interpret this result, that is, to understand what the world might be like if it is true that physical interactions between systems are non-local, or non-deterministic, or that physical systems do not possess definite properties. This line of thought found its climax in program of "experimental metaphysics" that developed after the violation of Bell-type inequalities was observed. Experimental metaphysics consists in deriving metaphysical conclusions from the Bell-type experimental results. The mainstream interpretation of this work is that Bell-type experiments force us to accept the existence of some form non-local and perhaps non-causal processes at the ontological level.

In my dissertation, I assess to what extent philosophical investigation can help us decide

what the world is like on the basis of our best physical theories, from the point of view of the quantum domain and with an emphasis on Bell-type phenomena. My conclusions point to a more modest view on the possible achievements of philosophy of physics than the experimental metaphysics program would have us believe.

In the first part of my dissertation, I investigate what role philosophy of physics can legitimately hope to play in the development and evaluation of various accounts of quantum phenomena. I claim that it is not the role of philosophy of physics to impose criteria of acceptability on physical theories, in addition to coherence and empirical adequacy. By contrast, I take in my dissertation that the legitimate role of philosophy of physics is to clearly determine what is imposed by the phenomena and our best theories from what is a matter of preference on the basis of the structural analysis of the phenomena and theories.

In the second part of my dissertation, I turn to the more specific case of the interpretation of Bell-type theorems and Bell-type phenomena. I undertake a systematic examination of the mainstream interpretation, according to which Bell-type phenomena force us to accept a form of non-locality, but a form that we can consider benign because it is of a non-causal type. I show that the mainstream interpretation includes three claims, one about locality, another about causation and a last one about holism. I utilize rigorous theories of locality and causality in order to assess these three claims. On the one hand, the upshot of my analysis is that the claim about locality can be supported by a rigorous theory of locality. On the other hand, neither of the types of theories of probabilistic causation available can support the claims of the mainstream interpretation about causality when it is construed as a strong program of experimental metaphysics yielding conclusions about the ontology of the world. That said, weakened versions of the mainstream interpretation, those that do without conclusions about the ontology of the world, can be made compatible with both types of theories of probabilistic causality. In particular, the mainstream interpretation can be rigorously supported if its claims are restricted to the empirical level.

Introduction – Français

Un théorie physique consiste en un appareil mathématique, le formalisme, accompagné de certaines règles de correspondance minimales, qui indiquent comment appliquer le formalisme à certaines situations physiques, et donc comment user du formalisme pour faire des prédictions empiriques. Si les théorie physiques sont minimalement pourvues de règles de correspondance, elle ne sont cependant en général pas pourvues d'une interprétation. Interpréter une théorie consiste à dire à quoi le monde peut ressembler si cette théorie est vraie. Proposer une interprétation consiste essentiellement à proposer une ontologie fondamentale pour la théorie et à indiquer comment de cette ontologie surgit le monde tel qu'il nous apparaît. Une theorie interprétée est généralement supposée nous offrir une meilleure compréhension de la fabrique du monde – dans les limites du domaine de la théorie, au niveau fondamental comme au niveau des apparences.

Il se trouve que les théories physiques classiques peuvent être interprétées comme représentant le monde en termes d'intéractions causales à la fois locales et déterministes entre des systèmes characterisés par des propriétés bien définies. Il est bien connu que la mécanique quantique orthodoxe, qui se trouve être par ailleurs une de nos théories les mieux confirmées expérimentalement, ne peut pas être interprétée comme les théories classiques. Les théorèmes de type Bell, and les expériences qui y sont associées, ont aggravé la situation en montrant qu'une telle interprétation est littéralement impossible. Au début des années soixante, John Bell a en effet démontré que toute théorie représentant son domaine en termes classiques satisfait une série d'inégalités, les inégalités de Bell. L'expérience montre que les phénomènes quantiques violent ces inégalités. Par un simple modus tollens, il suit qu'aucune théorie qui décrit le monde en termes classiques ne peut être compatible avec tous les phénomènes quantiques.

Dans de telles circonstances, certains sont tenté d'adopter un point de vue quasi-intrumentaliste. Le formalisme de la théorie quantique est alors conçu comme représentant notre connaissance des phénomènes, et non les phénomènes eux-mêmes. De nombreux physiciens et philosophes ne se satisfont cependant pas de la voie quasi-instrumentaliste. Au lieu de cela, des physiciens ont développé des théories alternatives à la théorie quantique orthodoxe, théories dont les interprétations abandonnent au moins une des charactéristiques du cadre classique: localité, determinisme, propriétés définies. Les philosophes s'efforcent alors d'interpréter ces nouvelles théories, i.e. s'efforcent de comprendre à quoi peut ressembler le monde s'il est vrai que les interactions entre les systèmes physiques sont soit non-locales, soit non-déterministes, ou bien que ces systèmes physiques ne possèdent pas de propriétés définies. Cette ligne de pensée a atteint son apogée avec la "métaphysique expérimentale" qui s'est développée après que la violation des inégalités de Bell par l'expérience a été observée. Le programme de la métaphysique expérimentale est fondé sur l'idée que nous pouvons dériver des conclusions métaphysiques des expériences de type Bell. L'interprétation la plus commune, que j'appelle l'interprétation orthodoxe, est que les expériences de type Bell nous forcent à accepter certaines formes d'interactions non-locales, et peut-être non causales, au niveau fondamental.

Dans cette thèse, je m'attache à la question générale de savoir dans quelle mesure l'investigation philosophique de nos meilleures théories scientifiques et des expériences associées peut nous informer de ce que peut être le monde au niveau fondamental. J'étudie cette question en considérant le domaine quantique, et en particulier les phénomènes de type Bell. Mes conclusions vont dans le sens d'une vue plus modeste du rôle possible de la philosophie de la physique que celle que les programmes de métaphysique expérimentale suggèrent.

Dans la première partie, j'étudie le rôle que la philosophie de la physique peut légitimement jouer dans le développement et l'évaluation des différentes theories quantiques. Je commence par présenter brièvement ces théories et leurs interprétations: la théorie de Bohm, les théories des mondes multiples, et les théories du type GRW (Ghirardi, Rimini, Weber). J'explique que de ces théories, munies de leur interprétations, abandonnent chacune un des éléments du cadre classique présenté ci-dessus. Respectivement: la localité, la définition des propriétés, ou bien l'évolution déterministique. Etant donné que ces théories sont chacune cohérente et toutes empiriquement équivalentes (dans les limites actuelles d'expérimentation), il semble que nous ne possédons pas de critère de choix satisfaisant.

Certains ont avancé l'hypothèse qu'un tel critère de choix pourrait résider dans le degré de compréhensibilité de ces théories munies de leur interprétation. Je montre que ce n'est pas le cas dans le second chapitre de la thèse. Je prends pour point de départ l'analyse de la compréhension des théories par de Regt et Dieks. Je soutiens que leur analyse est incomplète. Je défends que la compréhensibilité d'une théorie interprétée est la combinaison de la compréhensibilité de la théorie et de celle de l'interprétation. J'utilise mon analyse de la compréhensibilité pour comparer deux théories interprétées concurrentes: la théorie de Bohm et les théories des mondes multiples. Je montre que comparer la relative compréhensibilité de ces dernières ne permet pas de faire notre choix entre les deux. Je soutiens de plus que la compréhensibilité est en général un critère de choix faible entre théories interprétées, ceci simplement parce que la compréhensibilité d'une théorie et/ou de son interprétation n'a rien à voir avec la vérité de cette théorie et/ou de son interprétation, ce qui, au moins idéalement, reste notre but quand nous forgeons des théories.

A la lumière de ces arguments, je soutiens que le rôle de la philosophie de la physique ne consiste pas à imposer des critères d'acceptabilité pour les théories physiques en plus des critères usuels de cohérence interne et d'adéquation empirique. Un rôle à la fois important et légitime que peut cependant jouer la philosophie de la physique consiste à distinguer clairement, pour nos théories et leurs interprétations, ce qui est imposé par les phénomènes et les théories de ce que relèvent de nos choix et préférences. Je soutiens qu'un tel travail de distinction peut se faire sur la base de l'analyse structurelle de nos théories et de nos modèles de données. Je conclus la première partie de cette thèse en soutenant qu'une version modeste de la conception sémantique des théories scientifiques fournit un outil approprié pour une telle analyse, du fait qu'elle définit clairement les notions de structure et de relation structurelle. Je défends la conception sémantique contre les critiques récentes qui lui ont été récemment adressées dans la literature. Ces critiques reposent sur l'idée que les modèles scientifiques ne sont pas des modèles logiques et vice versa. Je soutiens que ces critiques ne tiennent que sous une interprétation forte de la conception sémantique, une interprétation qui plus n'est pas fidèle au projet initial tel qu'il a été défini et développé par Suppes depuis les années soixante. Je propose une interprétation modeste de la conception sémantique, and montre que, sous cette interprétation modeste, la conception sémantique est un programme de recherche non seulement tenable mais encore prometteur pour l'analyse des théories scientifiques.

Dans la seconde partie de la thèse, j'étudie spécifiquement le cas de l'interprétation des théorèmes de Bell et des expériences associées. J'évalue systématiquement l'interprétation orthodoxe selon laquelle les résultats de ces expériences nous forcent à accepter une forme "bénigne" de non-localité, "bénigne" parce que de type non-causal. Cette interprétation repose sur une distinction que Jarrett et Shimony ont faite entre deux notions de localité – indépendence vis à vis des résultats vs. indépendence vis à vis des paramètres, OI et PI, respectivement (pour *Outcome independence* et *Parameter Independence*) – qui conjointement impliquent la notion de localité qui permet la dérivation des inégalités de Bell. Il suit de cette analyse que les phémonènes de type Bell ne nous forcent à abandonner qu'une seule de ces formes de localité. Jarrett et Shimony soutiennent que les violations de OI, quand PI est respectée, correspondent à une forme bénigne de non-localité. Cette forme bénigne de non-localité a été subséquemment interprétée en termes d'interactions non-causales indicatrice d'une forme de holisme.

Dans ma thèse, j'entrepends un examen systématique de l'interprétation orthodoxe. Dans le chapitre 4, je présente en détails l'interprétation orthodoxe et ma stratégie d'analyse pour les chapitres à suivre. Je commence par montrer que l'interprétation orthodoxe comprends trois thèses différentes. La première thèse porte sur la question de la localité: les violations PI indiquent l'existence d'un processus non-local sous-jacent, mais non les violations de OI. La seconde thèse porte sur la question de la causalité: les violations de PI indiquent l'existence d'un lien causal sous-jacent, mais non les violations de OI. La troisième thèse consiste à interpréter ce lien non-causal en termes de holisme. Cette dernière thèse ne peut tenir que si la seconde tient également.

Une difficulté surgit quand on tente d'évaluer ces trois thèses du fait que, dans la littérature, la localité et la causalité sont charactérisées en termes de restrictions sur les distributions de probabilités conditionnelles. Ceci provient probablement de l'analyse de la causalité en termes de probabilités conditionnelles par Reichenbach et d'une association douteuse entre processus causaux et processus locaux. Il n'est pourtant pas clair que la théorie des probabilités conditionnelles à elle seule constitue un cadre adéquat pour discuter les questions de localité et de causalité. Ma stratégie dans cette thèse est de fournir des cadres théoriques rigoureux de discussion, pour la localité et pour la causalité, afin d'évaluer rigoureusement dans quelle mesure l'interprétation orthodoxe peut être soutenue.

Avant de fournir ma propre analyse de l'interprétation orthodoxe cependant, je dédis un chapitre entier aux travaux de Arthur Fine sur le sujet. Ceci parce que Fine a critiqué l'interprétation orthodoxe de façon constante. Je distingue trois stratégies d'argumentation utilisées par Fine, et je montre qu'aucune de ces stratégies n'est concluante. L'une de ces stratégies est de soutenir qu'il existe une prémisse cachée dans les dérivations usuelles des inégalités de Bell, à savoir la définition de certaines probabilités jointes. Si cela est vrai, alors tout ce que la violation des inégalités de Bell nous contraint de conclure est que le monde ne peut pas être représenté par des théories dans lesquelles ces probabilités jointes sont bien définies. Cette première stratégie d'argumentation échoue du fait qu'elle ne s'applique qu'à une classe restreinte de modèles depuis lesquels des inégalités de Bell peuvent être dérivées, à savoir les modèles non-contextuels, tandis que des modéles contextuels restent viables.

La seconde ligne d'argumentation de Fine est de faire usage de ses modèles en prisme, qui sont des modèles locaux pour les expériences de type Bell. Fine tente d'utiliser l'existence de ces modèles comme une preuve de facto que la localité n'est pas nécéssairement violée par tous les modèles des probabilités quantiques. Cet argument ne tient pas parce que les modèles en prisme sont certes des modèles des expériences actuelles mais non pas des modèles de la mécanique quantique elle-même. Dans les modèles en prisme, certains systèmes quantiques ne répondent pas à certains contextes de mesure. Ceci contredit la mécanique quantique, qui prédit que les systèmes quantiques répondent à tous types de mesures.

La dernière stratégie mise en place par Fine consiste à soutenir que l'ensemble d'hypothèses utilisé pour les dérivations des inégalités de Bell est inconsistent, de sorte que la contradiction obtenue à la fin ne devrait pas surprendre. Je montre que ceci n'est vrai que sous une hypothèse forte concernant le status ontologique des probabilités. Cette hypothèse est controversée et Fine ne fournit pas d'argument en sa faveur, de sorte que cette dernière stratégie échoue également. Ma conclusion est donc que les arguments développés par Fine contre l'interprétation orthodoxe ne sont pas concluants.

Je développe ma propre analyse de l'interprétation orthodoxe dans les deux derniers chapitres. Je commence par la question de la localité. Logiquement, une définition de la localité devrait être intimement liée à la structure spatiotemporelle dans laquelle les événements considérés sont inclus. Prima facie, des restrictions sur les distributions de probabilités conditionnelles n'indiquent pas à elles seules comment les événements sont inclus dans une structure spatiotemporelle. Je propose un cadre spatiotemporel rigoureux dans lequel j'étudie la question de savoir si PI et OI (indépendence vis à vis des paramètres et indépendence vis à vis des résultats, respectivement) sont des conditions de localité. Je montre que la version stochastique de la localité Einsteinienne appliquée aux situations de type Bell implique PI mais non OI. Il résulte de cela que, dans le cadre théorique choisi, l'interprétation orthodoxe est tenable concernant les thèses sur la localité.

Afin d'évaluer les thèses de l'interprétation orthodoxe concernant la causalité, je fais appel à deux théories de causalité probabilistes, à savoir, les théories manipulabilistes (MTPC) et les théories spatiotemporelles (STPC) de la causalité. J'utilise ces théories afin de déterminer si les violations de PI et/ou de OI indiquent des liens causaux sous jacents. Je montre que, sous une application stricte des MTPC, un argument satisfaisant peut être développé en faveur de la thèse selon laquelle les violations de PI, mais non pas les violations de OI (tandis que PI est respectée), correspondent à l'existence d'un lien causal, mais ceci uniquement au niveau empirique. Tout ce que l'on peut conclure de l'application des MTPC aux phénomènes de Bell est que, contrairement aux violations de PI, les violations de OI, quand PI est respectée, n'indiquent pas de lien causal qui ait de conséquences empiriques. Ceci est loin d'être suffisant pour la métaphysical expérimentale, mais pourrait cependant contenter de nombreux physiciens and philosophes.

Les STPC fournissent une théorie métaphysique de la causalité et pourraient sembler prometteuses. Je montre cependant que sous une application stricte des STPC, l'interprétation orthodoxe n'est pas tenable. Ceci dit, je montre que, considérant une version améliorée des STPC, il devient possible de défendre une version faible de l'interprétation orthodoxe, selon laquelle les violations de PI sont indicatrices d'un lien causal sous-jacent, tandis que les violations de OI indiquent seulement que nous devrions considérer les deux sous-systèmes comme des parties d'un unique système.

Le résultat de mon analyse est qu'aucune des théories de la causalité considérées ne permet de soutenir la version forte de l'interprétation orthodoxe comprise comme métaphysique expérimentale. Considérant que Butterfield a montré que cette interprétation ne tient pas non plus dans les théories contrefactuelles de la causalité telle que proposées par Lewis, le résultat est que la version forte de l'interprétation orthodoxe ne tient dans aucune des théories de la causalité actuelles.Dans les deux cas cependant, une version faible peut être défendue. En particulier, un résultat important de cette thèse est que l'interprétation orthodoxe peut être défendue rigoureusement au niveau empirique.

Introduction – English

A physical theory consists in a mathematical apparatus, the so-called formalism, along with some basic correspondence rules that indicate how to apply the formalism to certain physical situations, thus allowing us to make empirical predictions. To provide a physical theory with an interpretation is to give an account of what the world could be like if the theory was true. An interpretation primarily consists in an articulation of the fundamental ontology associated with the physical theory, as well as an indication of how the appearances emerge from the proposed fundamental ontology. An interpreted physical theory is commonly taken to give us some understanding of what the physical world might be like and how it appears to us the way it does, within the domain of application of the theory.

It so happens that classical physical theories can be interpreted as a representation of local, deterministic, causal interactions between systems with definite properties. Orthodox quantum theory, which happens to be one of our most experimentally well-confirmed theories, is notoriously resistant to being interpreted in terms of the above framework. Bell-type theorems and Bell-type experiments have made such an interpretation almost impossible. In the early sixties, John Bell demonstrated that any theory that represents its domain in terms of the above framework satisfies a set of inequalities, the so-called Bell inequalities. Experiments on quantum phenomena violate Bell-type inequalities. By a simple modus tollens, the upshot is that no theory that includes all the elements of the above framework can give an account of all quantum phenomena.

Given this situation, one response is to adopt a quasi-instrumentalist point of view. The formalism of quantum theory can be taken to represent our knowledge of the phenomena, rather than the phenomena. Many physicists and philosophers do not content themselves such a view. Physicists have developed alternative accounts of quantum phenomena, in which at least one element of the framework above is abandoned. Philosophers have been trying to interpret this result, that is, to understand what the world might be like if it is true that physical interactions between systems are non-local, or non-deterministic, or physical systems do not possess definite properties. This line of thought has found its climax in the program of "experimental metaphysics" that developed after the violation of Bell-type inequalities was observed. Experimental metaphysics consists in deriving metaphysical conclusions from the Bell-type experimental results. The mainstream interpretation is that Bell-type experiments force us to accept the existence of some form non-local and perhaps non-causal processes at the ontological level.

In my dissertation, I assess to what extent philosophical investigation can help us decide what the world is like on the basis of our best physical theories, from the point of view of the quantum domain and with an emphasis on Bell-type phenomena. My conclusions point to a more modest view on the possible achievements of philosophy of physics than the experimental metaphysics program would lead us to believe.

In Part I of my dissertation, I investigate what role philosophy of physics can legitimately hope to play in the development and evaluation of various accounts of the quantum phenomena. In the first chapter, I begin by briefly presenting these accounts: Many Worlds Theories, Bohm-type theories and GRW collapse theories, explaining how each gives up one of the elements of the framework presented above: definite properties, locality, or deterministic evolution, respectively. Given the internal coherence of, and the empirical equivalence (up to our current experimental capabilities) of these accounts, we do not seem to possess good reasons to favor one of these interpreted theories over others.

Some have claimed that the degree to which an interpreted theory is understandable could be used as a good criterion of choice between interpreted theories. In Chapter 2, I argue against such a view. I begin with the de Regt-Dieks analysis of understandability of a theory. I argue that their analysis is incomplete. I propose that the understandability of an interpreted theory is the combination of the understandability of its theory as well as the understandability of its interpretation. I use my extended analysis of understandability to compare two competing interpreted theories: Bohm's theory and Many-World theory. I argue that understandability does not provide a means to decide between the two. Moreover, I argue that understandability is in general a weak criterion of choice for interpreted theories because it has nothing to do with the truth of a theory, which remains, even ideally, our goal when constructing theories.

In the light of the foregoing, I claim that it is not the role of philosophy of physics to impose criteria of acceptability on physical theories, in addition to coherence and empirical adequacy. In particular, it is not the role of philosophy of physics to impose such criteria on the basis of contingent features of human cognition. By contrast, I take in my dissertation that the legitimate role of philosophy of physics is to clearly determine what is imposed by the phenomena and our best theories from what is a matter of preference on the basis of the structural analysis of the phenomena and theories. I argue in Chapter 3 that a modest version of the semantic view of scientific theories, in so far as it offers clear definitions of structure and of structural relationship, constitutes the appropriate tool for such structural analysis. In defense of the semantic view, I address the criticisms recently leveled against the semantic view in the literature, that scientific models are not logical models and vice versa. I argue that such criticisms rely on a strong interpretation of the semantic view, an interpretation that is arguably unfaithful to the initial project, as defined and developed by Suppes since the sixties. I propose a modest interpretation of the semantic view, and show that the modest version of the semantic view is both tenable and a promising research program for the analysis of scientific theories.

In the second part of my dissertation, I turn to the more specific case of the interpretation of Bell-type theorems and Bell-type phenomena. I assess the mainstream interpretation according to which Bell-type phenomena force us to accept a form of non-locality, but a form that we can consider benign because it is of a non-causal type. The mainstream interpretation relies on a distinction that Jarrett and Shimony have drawn between two notions of locality – outcome independence and parameter independence – which together imply the notion of locality from which Bell-type inequalities are derivable. Given this analysis of locality, Bell-type phenomena force us to give up only one of these two forms of locality. Jarrett and Shimony claim that failure of outcome independence, when parameter independence holds, corresponds to a benign form of non-locality. Such a benign form of non-locality has been further interpreted in terms of non-causal interaction indicating some sort of holism.

In my dissertation, I undertake a systematic examination of the mainstream interpretation. In Chapter 4, I present the mainstream interpretation and my strategy of analysis. I begin by arguing that the mainstream interpretation includes three different claims. A first claim concerns the issue of locality: parameter independence is a locality condition, while failure of outcome independence is not (PI-LOC). A second claim concerns the issue of causation: failure of parameter independence is indicative of a causal underlying process, while failure of outcome independence is not (¬PI-CAUS). The third claim consists in interpreting the failure of outcome independence in this situation as a form of holism (¬OI-HOL). Hence, the third claim cannot stand if the second does not.

When it comes to assess these three claims, a difficulty arises from the fact that, in the Bell literature, locality and causation are solely characterized in terms of restrictions on conditional probability distributions over events. Such a focus on probability distributions seems to be a remnant of Reichenbach's analysis of causation in terms of probability distributions and a challengeable association of local and causal processes. Additionally, it is not clear that the framework of conditional probabilities (alone) is appropriate for dealing with the issues of locality or causality. My strategy in my dissertation is to utilize rigorous theories of locality and probabilistic causation in order to assess to what extent the mainstream interpretation can be rigorously supported.

Before I give my own analysis however, I devote an entire chapter to the work Bell-type theorems and Bell-type phenomena by Arthur Fine (Chapter 5). This is because Fine has been constantly challenging the mainstream interpretation for the last twenty-five years. I distinguish between three different strategies in Fine uses to argue his point. I show that none of these strategies is successful.

One of his strategies is to claim that there exists a hidden assumption in the framework from which Bell-type inequalities are usually derived, namely, the definition of some joint probabilities. If true, then all we had to conclude from the violation of Bell-type inequalities is that the world cannot be represented by theories in which these joint probabilities are well defined. This first strategy of argument fails because it concerns only a restricted class of frameworks for the derivation of Bell-type inequalities, namely, non-contextual frameworks, and contextual frameworks are possible.

A second line of argument makes use of Fine's famous Prism Models, which are local models for Bell-type experiments. Fine hopes to use the existence of his models as a de facto proof that locality does not have to fail in models of quantum phenomena. In short, this argument fails because Prism Models are models of the actual experiments and not models of quantum mechanics. Within Prism Models, some quantum systems will not respond to some measurements. This is in contradiction with quantum theory, which predicts that quantum systems will always respond to measurements.

Fine's final strategy is to claim that the derivations of Bell-type inequalities starts with an inconsistent set of assumptions, so that the final contradiction is not surprising. I show that this is only true under a strong assumption about the ontological status of probabilities. Such an assumption is controversial and unsupported, Fine's last strategy of argument is unsuccessful. At the end of the day, Fine's arguments to dismiss the importance of Bell-type theorems and Bell-type phenomena are not conclusive: Bell-type theorems and phenomena are intriguing and deserve our attention.

In the last two chapters of the dissertation, I give my own analysis of the mainstream interpretation. Chapter 6 deals with the issue of (PI-LOC). Arguably, a definition of locality should be intimately related to the spacetime structure that events are embedded in. Prima facie, restrictions on conditional probability distributions do not alone indicate how events are embedded within a spacetime structure. I propose a rigorous spacetime framework in which I assess whether or not parameter and outcome independence are locality conditions. I show that a rigorous argument can be given to the effect that the stochastic version of Einstein Locality entails parameter independence but not outcome independence. The upshot is that a rigorous argument can be made in favor of (PI-LOC).

In order to examine the causal relationships in Bell-type experiments I utilize two theories of probabilistic causation; namely, manipulability theories of causation (MTPC) and spacetime theories of causation (STPC). I use these theories to determine whether or not failure of outcome independence and parameter independence are indicative of underlying causal interactions.

I show that under a strict application of MTPC, a good argument can be made in favor of (¬PI-CAUS), but only at the empirical level. Failure of parameter independence is indicative of a causal link with empirical consequences, but failure of outcome independence is not indicative of any causal link which can manifest itself as the empirical level. So MTPC support the idea of' "peaceful coexistence" between models of quantum phenomena in which only outcome independence is violated while parameter independence holds and Relativity (if Relativity is taken to forbid superluminal causation) at the empirical level, but leave open what the world is like behind the phenomena. This is far from supporting any kind of "experimental metaphysics" but still could content many physicists and philosophers.

STPC provide a more metaphysically oriented account of causation. However, we show that under a strict application of STPC, the mainstream interpretation fails. That said, I argue that under a refined version of the spacetime theories of probabilistic causation, some plausibility arguments can be made in favor of a weakened version of the traditional interpretation, according to which failure of parameter independence is to be regarded as indicative of a causal interaction between the parameters and the outcomes, while failure of outcome independence indicates that we should regard the two subsystems as parts of a single unified system.

Given that Butterfield has proved elsewhere that the mainstream interpretation fails within the counterfactual theories of probabilistic causation à la Lewis, the upshot of my analysis is thus that neither of types of theories of probabilistic causation available can support the mainstream interpretation construed as a strong program of experimental metaphysics. That said, it is shown weakened versions of (¬PI-CAUS) and (¬OI-HOL) can be made compatible with some theories of probabilistic causation. In particular, an important result of this dissertation is that the mainstream interpretation can be rigorously supported if its claims are restricted to the empirical level.

Part I

Understanding quantum phenomena

Chapter 1

Three ways of understanding quantum phenomena

1.1 Introduction

Most of us take it that our best physical theories give us some understanding of how the world works and of what it is made of. This of course, is assuming that we can understand these theories.[1] Concerning quantum phenomena, however, it is a common claim, say, a claim that everybody has heard from unphilosophically-minded physics professors in undergraduate classes, that quantum phenomena are not really understandable, even if quantum physics is one of the most empirically well confirmed and most successful theories. The contrast class here is classical phenomena which supposed to be understandable. Further, it is usually added that this situation is not a problem because all we ask from quantum theory is that it makes correct predictions about quantum phenomena, not that it gives us any kind of fundamental account of the quantum domain. This is to say that standard quantum mechanics equipped with the so-called orthodox interpretation is not a fundamental theory.

[1] As we shall explain in Chapter 2, there are various ways in which an interpreted theory is understandable. We do not need such an analysis in this chapter.

Many philosophers and physicists however do not content themselves with such a quasi-instrumentalist view. Physicists and philosophers have developed interpretations for quantum mechanics as well as alternative theories for the quantum domain and interpretations for those theories. Providing an interpretation of theory is to indicate what the world might be like if the theory is true.

Many think it a legitimate role for philosophers to provide criteria of choice between various (coherent and empirically adequate) accounts that would make it possible to choose between the various ontologies associated with these theories. In this Part of the dissertation, we shall deny this role for philosophers. We shall argue that a modest but legitimate role for philosophers of physics is to determine what is imposed by the phenomena and our best theories from what are matters of preference.

In order to do so, it is useful to lay the groundwork for alternative theories of quantum physics and their interpretations. This is the burden of this chapter. In Section 1.2, we begin by demonstrating that orthodox quantum mechanics does not qualify as a good candidate for a fundamental theory because of the measurement problem. That said, orthodox quantum mechanics is but one of many theories that we have today for the quantum domain. There are other coherent and empirically adequate alternative accounts of quantum phenomena, and we shall briefly describe three of them. Each of them give a different solution to the measurement problem. We shall describe these three alternatives in Section 1.3. Section 1.4 provides an outline for this first part of this dissertation.

1.2 Orthodox quantum mechanics and the measurement problem

1.2.1 The orthodox interpretation

From its inception, the interpretation of quantum mechanics has been the object of debate. It is well known that Einstein and Shrödinger opposed the so-called “Copenhagen interpretation”[2] usually associated with Bohr and Heisenberg. Further, it is well known that the debate raged within the Bohr Institute. At least as usually understood, followers of Bohr defended a roughly neo-kantian view and Heisenberg defended a roughly positivist view.[3] von Neumann provided a rigorous axiomatization of the theory that was and has been accepted by most working physicists.[4] This framework provided a minimal interpretation is what we shall call “orthodox quantum mechanics”.[5]

In orthodox quantum mechanics the pure states of a system are represented by unit-length vectors (up to an overall phase factor) in the appropriate Hilbert space. The physical properties of a system that one might observe are represented by projection operators in the Hilbert space – the so-called observables. In general, systems do not possess definite properties for all their observables: there is a special rule for ascribing properties to a system, namely, the Eigenstate - Eigenvalue rule (E-E rule). The E-E rule tells us that a system has the property corresponding to a given eigenvalue of a given observable if and only if its quantum state is an eigenstate of the observable (more generally, if the corresponding eigenprojection is given probability 1 by the quantum state). It is indeterminate otherwise.

[2] The name comes from the Institute for Theoretical Physics which Niels Bohr was the director in the twenties – it was later renamed the Niels Bohr Institute. Under his direction, the Copenhagen Institute was a major center of research for the formulation, development and interpretation of quantum mechanics.

[3] It is not the place here to give a precise account of these debates. For more details, see [76] and references therein.

[4] [132].

[5] Note that we are not attempting to describe the so-called “Copenhagen Interpretation” of quantum mechanics.

Concerning the dynamics, two different kinds of interactions must be distinguished, for which two different dynamical laws apply. As long as no measurement is performed on a system, the dynamics is given by the Schrödinger equation. Such evolution is linear and deterministic; it depends only on the Hamiltonian that describes the system. When a measurement is made on the system, the dynamics is governed by the Collapse Postulate. It states that upon measurement the state of the system collapses to one of the eigenstates of the observable that was measured. Such evolution is non-linear and indeterministic.

So, the theory does not indicate which state a system is going to collapse to upon measurement. In other words, the theory does not tell which definite property is going to be observed after a measurement. That said, the theory gives the probabilities associated with the various possible outcomes – by the Born Rule. The Born rule states that when a state is written in the measurement basis, the probabilities of certain measurement results are given by the square of the complex coefficients of the basis vectors corresponding to those results.

It should be stressed that the projection postulate and the Born rule are uncontroversial as practical rules. Orthodox quantum mechanics is one of our best empirically confirmed theories. Difficulties only emerge if one tries to interpret orthodox quantum mechanics as a fundamental theory.

1.2.2 The measurement problem

In this subsection, we explain why orthodox quantum mechanics cannot be taken to be a fundamental theory. There are at least two necessary criteria for a theory to be considered as a candidate for a fundamental theory. A first criterion is that the theory gives a dynamical account of all physical processes and interactions in its domain. In the case of quantum mechanics, this includes measurement interactions. A second criterion is that the theory saves the phenomena. This is simply to say that a theory can be considered fundamental only if it rightly predicts what we observe. No claim is made here that these criteria are sufficient for a theory to be considered fundamental. These two criteria are minimal desiderata for a theory to be a candidate fundamental theory, and a fortiori for a theory to be considered as a (structurally, approximately etc., according to your own preferences in the debate over scientific realism) true description of the world.

Orthodox quantum mechanics as characterized above cannot be a fundamental theory because it does not provide a dynamical account of measurement interactions in the sense that there are neither necessary nor sufficient conditions for defining exactly what a measurement is, and hence, for the application of the two dynamics to a system. Moreover there is no clear physical justification for why measurements should be treated as a special kind of interaction. In Bell's evocative words:

> It would seem that the theory is exclusively concerned about 'results of measurement' and has nothing to say about anything else. What exactly qualifies some physical systems to play the role of 'measurer'? Was the wave function waiting to jump for thousands of millions of years until a single celled living creature appeared? Or did it have to wait a little longer, for some better qualified system...with a PhD? If the theory is to apply to anything but highly idealized laboratory operations, are we not obliged to admit that more or less 'measurement like' processes are going on more or less all the time, more or less everywhere? Do we not have jumping all the time?[6]

A natural move is to include measurement processes in the domain which is governed by the Schrödinger equation. This is how the measurement problem emerges. Due to the linearity of the Schrödinger equation, an interaction between two systems, one of which is in a superposed state, results in a state of superposition for both systems. Importantly, the pointer observable for the measurement device will be in a superposition state. It is uncontroversial that there is no ignorance interpretation for quantum superpositions whereby the

[6] [10].

system could be viewed as really pointing to one particular outcome, though the formalism does not indicate which. This is unfortunate because this is simply not what we seem to observe at the macroscopic level. Experiments appear to have determinate outcomes. Just adopting the Schrödinger evolution is thus not enough to show how one can recover the phenomena.

Given this situation, one can attempt to resolve the problem of recovering the appearances by utilizing the theory of decoherence. Decoherence theory provides a dynamical account for the suppression of interference, at a certain level of description, through spontaneous interaction with the environment, which is in agreement with Schrödinger evolution. It is rather uncontroversial that decoherence is relevant to why observable properties appear classical to us. It is, however, also uncontroversial that appealing to the theory of decoherence does not solve the measurement problem at the fundamental level.[7] As mentioned above, it does not provide any new dynamics additional to the Schrödinger equation. Hence, by linearity of the Schrödinger evolution, any enlarged system, which includes the apparatus and the environment, ends up in a superposition state. Thus, even if decoherence theory accounts for classical behavior at the level of the components, it does not suppress the problem that the different components are superposed. By exactly the same argument as the one used in stating the original measurement problem, a macroscopic physical system is represented by a superposition state by the theory, a state for which there is no clear and uncontroversial interpretation. To push the argument a little bit further, decoherence makes things even worse since not only the apparatus, but also the environment and finally the entire universe can end up being represented by a superposition state.

From the above, one can see that the measurement problem arises because orthodox quantum mechanics holds the three following claims:

1. the wave function gives a complete description of any physical systems;

[7] For details on this, see for example [140] and [9].

2. physical systems have definite properties given by the E-E rule;

3. the wave function evolves linearly and deterministically.

The measurement problem shows that no fundamental theory can include these three claims together. Any theory of the quantum domain has to give up at least one of the above claims in order to qualify as a candidate for a fundamental theory.

There are three different accounts of quantum phenomena that can be distinguished from one another by which feature (1.,2., or 3.) they give up.

- Bohm's theory gives up 1., the idea that the wave function is a complete account of physical systems;

- The Many-Worlds theories give up 2., the idea that physical systems have definite properties at the fundamental level;

- GRW theories[8] give up 3., the idea that the wave function evolves linearly and deterministically.

We shall briefly describe these three ways of understanding quantum phenomena in the following section.

1.3 Three ways of understanding quantum phenomena

Arguably, all three theories mentioned above solve the measurement problem. They give three different accounts which are both coherent and empirically equivalent, while suggesting very different worldviews.

[8] GRW stands for Ghirardi, Rimini and Weber

1.3.1 Bohm's theory

In Bohm's theory, one rejects the idea that the wave function be a complete description of physical systems. Bohm's theory supposes that quantum systems are particles and the description of quantum systems by means of a wave function is completed by the specification of configuration of the particles that compose the system, a specification of particle positions. These particles evolve according to the guiding equation, which relates time evolution of the configuration of the particles to the wave function. The wave function as well as the configuration of the particles evolve deterministically. Unfortunately, the theory has as a consequence that we can never determine the configuration of particles with accuracy greater than standard quantum mechanics. There is always uncertainty associated with the configuration of the particles. This is how probabilities arise in the theory. In Bohm's theory, all properties of a system, save for the configuration of a system and properties which are functions of the configuration of a system, are dispositional properties. The dispositional properties of a system are reducible to the categorical properties of the positions of the particles that compose a system given a specific measurement context. So, measurement results always supervene on the configuration of a system. Because we can never be sure of the actual configuration of a system, we can never know which outcome will occur in measurement situations, despite the fact that systems evolve deterministically. It so happens that Bohm's theory is empirically adequate and recovers the heretofore successful predictions of the standard theory.[9]

In short then, Bohm's theory gives us a way to understand the quantum domain as constituted of particles with definite positions and in deterministic motion. One can understand Bell's enthusiasm upon his discovery of the existence of such a theory:

> But in 1952 I saw the impossible done. It was in papers by David Bohm.

[9] According to several authors, this is only an "equillibrium" feature of the theory, and in principle one could have empirical predictions distinct from those of quantum mechanics.

> David Bohm showed explicitly how parameters could indeed be introduced, into nonrelativistic wave mechanics, with the help of which the indeterministic description could be transformed into a deterministic one. More importantly, in my opinion, the subjectivity of the orthodox version, the necessary reference to the 'observer', could be eliminated.[...]
>
> But why then had Born not told me of this 'pilot wave'? If only to point out what was wrong with it? Why did von Neumann not consider it? More extraordinarily, why did people go on producing 'impossibility proofs' after 1952, and as recently as 1978? When even Pauli, Rosenberg, and Heisenberg, could produce no more devastating criticism of Bohm's version than to brand it as 'metaphysical' and 'ideological'? Why is the pilot wave picture ignored in textbooks? Should it not be taught, not as the only way, but as an antidote to the prevailing complacency? To show that vagueness, subjectivity, and indeterminism, are not forced on us by experimental facts, but by deliberate theoretical choice?[10]

Advocates of Bohm-type theories say that the great advantage of these theories is that they make clear *what quantum theory is about.* Indeed, Bohm-type theories are associated with an ontology of particles in deterministic movement. Now, of course, the classical picture of the world is not completely recovered. The most important departure from the classical picture is that Bohm-type theories are highly non-local: the evolution of any particle depends on the position of all other particles. Also, a serious drawback is that there is no relativistic formulation available for Bohm-type theories. These are the tradeoffs involved in retaining a particle ontology in the quantum domain.

[10] From Bell, [14, p. 160]

1.3.2 Many Worlds

The central idea of the Many-Worlds interpretations of no-collapse quantum mechanics is to take the wave function seriously. More precisely, it is to accept the wave function and its deterministic evolution as a complete description of any physical system: the wave function is universal and never collapses. In a Many-Worlds interpretation one has to reject the idea that physical systems possess definite properties. It will be recalled that in general, quantum systems are described as having indeterminate properties. The orthodox interpretation via the E-E Rule ensures that some properties have definite values. By contrast, the Many-Worlds interpretations tell us that the world is made of physical systems with no determinate properties at the fundamental level.

Given this fundamental ontology, how are we to recover the appearance of determinate properties? The crucial idea of Many-Worlds theories includes observers and measuring devices in our description of any physical situation. The observer is a physical system just as any other one. After a measurement interaction, the state of an observer and the measurement device will be in a superposition state and are entangled: they are no longer capable of independent description. This state is interpreted as a realization of all possible measurement outcomes and an observation of all measurement outcomes as well. Metaphorically, we can think of the world splitting during measurement interactions. In each "world" an observer will appear to see a definite outcome for measurement interactions.

Many-Worlds theories tell us that the world is made of physical systems that lack definite properties at the fundamental level, even though appearances of definite properties emerge at the level of the observer. They give an account of the quantum domain which is local, deterministic and compatible with Relativity(although this is a subtle point). A typical objection to Many-Worlds theories is to suggest that the ontology of worlds is "superfluous". Exactly what counts as superfluous structure is typically in the eye of the beholder, and is a weak objection to the Many-Worlds theories. There is also a problem understanding probabilities in this case. There has been a great deal of interesting work on the matter recently, and the problem of probabilities does not seem incapable of a solution. Hence, Many-Worlds theories seem to be reasonable contenders for fundamental theories.

1.3.3 Collapse Theories

Collapse Theories à la Ghirardi, Rimini and Weber are the third account of quantum phenomena that qualifies as a candidate for a fundamental theory. Collapse Theories give up on the idea of a deterministic and linear evolution of quantum system. The crucial motivation is to take the collapse of the wave function upon measurement as a real physical process. Collapse theories recover definiteness of measurement outcomes by postulating spontaneous collapses of the wave function. The upshot is a world view which is fundamentally indeterministic. Collapse theories replace the deterministic evolution of orthodox quantum mechanics with stochastic evolution where:

1. In clear cut cases of measurement, predictions of the theory approximate those that would be made using the projection postulate on the standard theory;

2. In clear cases of no measurement, dynamics of systems are roughly equivalent to those given by the Shrödinger equation.

Roughly, the quantum mechanical state of a system follows the Schrödinger equation except that it has a probability distribution for *spontaneous* collapse (independently of whether it is measured or not). One can fix such probability distributions together with other factors so that the two conditions above are satisfied. Though not strictly empirically equivalent to standard quantum mechanics, Collapse theories make predictions equivalent to orthodox quantum theory up to our current experimental capacities. Also, it appears that Collapse theories can be given a relativistic formulation. One major drawback of these theories is that energy isn't suitably conserved. That said, these theories are in their infancy, and have no objections that seem insurmountable to date.

1.4 Outline of Part I

In this Part of the dissertation, we will use the above discussion as a starting point to argue that philosophers of physics should concern themselves with determining, in theories and interpretations, what is imposed by the phenomena or our best theories from what are matters of preference. In Chapter 2 we argue that philosophical investigation can give not provide us with criteria for theory choice, besides coherence and empirical adequacy, for theory choice.

In order to distinguish, in theories and interpretations, what is imposed by the phenomena or our best theories from what are matters of preference, one has to have in hand a clear notion of what a theory is and how theories are related to one another. In Chapter 3 we will argue that the Semantic view of theories is the appropriate tool for the job, defending it from recent criticisms. With this framework in hand, we shall turn to the investigation of the interpretation of Bell-type theorems and Bell-type phenomena in Part II.

Chapter 2

How understanding matters – or not.

The aim of this chapter is to show that understandability is not a basis for choosing between Bohm's theory and the Many-Worlds interpretation of standard quantum mechanics. Advocates on both side assert that their preferred account is more understandable than the other. On that score, they are both right. The seeming inconsistency involved in this claim is dissolved when one realizes that they employ different notions of understandability. Moreover, understandability, on either notion, is not an overriding criterion of choice between competing accounts if our aim in developing physical theories is truth.

It is well known that there are several ways to account for quantum phenomena. Bohm's theory (in the equilibrium case) and the Many-Worlds interpretation of standard quantum mechanics (no-collapse, no extra values) seem to be among the most competitive on the market. As it stands, we do not seem to possess any good criterion to choose between them. There are several criteria on the basis of which we usually choose between theoretical accounts of a set of physical phenomena. Internal consistency, explanatory and unificatory power are among such criteria. Arguably though, the most important is empirical adequacy. The main problem, when it comes to choosing between Bohm's theory and the Many-Worlds interpretation of standard quantum mechanics (SQM) is that they are empirically equivalent.

Defendants on each side try to argue for the superiority of their account of quantum phenomena. Some have defended accounts on the basis of its higher degree of "understandability". Such a defense begins with the following premises:

1. One of the accounts is more "understandable" than the other;

2. Understandability is a good basis for a criterion of choice.

They conclude that their favored interpretation is the one we ought to adopt. In this paper, I would like to show that this is a poor argument. Sections 1 and 2 highlight problems with premise 1. Section 3 highlights problems with premise 2.

Claims about the higher degree of understandability of each account over the other typically are along the following lines. Bohmians claim that Bohm's theory is more understandable because it has a "clear ontology". To this however, Many-Worlders have answered that their account is more understandable because it is closer to the standard formalism and has a "minimal" ontology. We would like to show that such claims involve two different meanings of understandability, and the Bohmians and Many-Worlders are talking past one another. With the analysis in hand, we shall assess to what extent Bohm's theory and the Many-Worlds interpretation can respectively be said to be "more understandable" than the other.

A notion of understandability is famously very difficult to define. The analytic tradition since Hempel ([66, 413]) has rejected it from the domain of philosophy for want of an objective definition. It should be noted from the outset that we shall not attack the notion of understanding and understandability as psychological or subjective. We shall partly rely on the work of de Regt and Dieks in [33] who give a general account of scientific understanding. We hope to contribute to a completion of their analysis.

The confusion in the Bohm vs. Many-Worlds debate over understandability stems in part from the failure to recognize that Bohm's theory and the Many-Worlds interpretation are not competing interpretations of the same formalism but competing interpreted theories. We shall propose that the understandability of an interpreted theory consists in the combination of the understandability of its theory (T-understandability) and the understandability of its interpretation (I-understandability). We shall further claim that, in the Bohm vs. Many-Worlds debate, the sense of understandability that the Many-Worlder uses is the former, while Bohmians use the latter. We will argue that these notions of understandability do not provide a means deciding the Bohm vs. Many-Worlds debate over understandability. Moreover, we will argue that understandability, generally speaking, is a weak criterion of theory choice because it has nothing to do with the truth of a theory.

In Section 2.1, we will first introduce some terminological clarifications that are useful for getting a handle on what is at issue in the Bohm vs. Many-Worlds debate. We will proceed to define two different notions of understandability and discuss how they are related. In Section 2.2 we will discuss the understandability of Bohm's theory and the Many-Worlds theory in relation to the notions defined. In Section 2.3 we will argue that understandability is a poor criterion of theory choice.

2.1 Notions of Understandability

2.1.1 Terminological clarifications

One of the confusions that pervades the Bohm vs. Many-Worlds debate is to think of them as two competing interpretations for the same formalism. It should be noted right from the outset that Bohm's theory is not an interpretation of standard quantum mechanics because it utilizes a different formalism. The following clarifications in our basic terminology will prove useful for the upcoming analysis of understandability:

Physical theory – T – A physical theory consists primarily in a mathematical apparatus, the so-called formalism, F. That said, the formalism alone does not suffice to constitute a *physical* theory. The formalism comes with some basic correspondence rules, CR, which tell us how to apply the formalism to certain *physical* situations, thus allowing us to make some empirical predictions.

Thus, we have $T = F + CR$

We hold throughout that theories are chosen primarily for their empirical adequacy.[1]

Interpretation – I – Whereas physical theories usually come with some basic correspondence rules, they do not come with a complete interpretation. To provide a physical theory with an interpretation is to give a story of what the world could be like if the theory were true. An interpretation I is primarily an ontological framework for the formalism. An interpretation tells us what entities and what physical quantities correspond to the mathematical constructions of the formalism. A constraint on acceptable interpretations is that they are able to recover the appearance of the world to us, even when it may seem strikingly at odds with the proposed fundamental ontology.

Different interpretations of the same formalism are, by definition, empirically equivalent. We do not have definitive criteria for choosing between consistent interpretations which are compatible with all the empirical predictions of a theory.

Interpreted theory – IT – An interpreted theory is a physical theory provided with an interpretation. This is to say, an interpreted theory consists in 1. a formalism, 2. some basic correspondence rules, and 3. an interpretation.

Thus, we have: $IT = F + CR + I$.

Bohm's theory and the Many-Worlds interpretation, instead of being competing interpretations of the same theory, are really competing interpreted theories. They differ not only

[1] "Empirical adequacy" here is not intended in the sense of van Fraassen in [128]. We simply mean that the theory has made successful predictions in its intended domain of application.

as far as their ontological and epistemological assumptions are concerned, but also formally. In order to be clear at all times, I shall use the following acronyms in the remaining of the paper:

- BIT: the full package of Bohm's formalism, correspondence rules, plus a Bohmian type interpretation
- BT: Bohm's formalism and correspondence rules
- BI: a Bohmian interpretation
- MWIT: the full package of standard quantum mechanics (no collapse, no extra values) plus a Many-Worlds-type interpretation
- MWT: the physical theory used in MWIT, that is, standard quantum mechanics
- MWI: a Many-Worlds-type interpretation

These clarifications form the foundation of our analysis of notions of understandability and how they relate to Bohm's theory and of the Many-Worlds interpretation.

2.1.2 Two different notions of understandability

Our main point in this section is that the notion of understandability of physical theory, T-understandability, is different than the notion of understandability of an interpretation, I-understandability. We shall maintain that each of these two kinds of understandability seem to independently contribute to the understandability of the interpreted theory.

IT-understandability The understandability of an interpreted theory consists in the interpreted theory being both I-understandable and T-understandable.

The issue of how the understandability of the theory and of the interpretation respectively contribute to the understandability of the interpreted theory will be addressed later. We shall now try to contrast the two kinds of understandability.

Understandability of a physical theory

Concerning what it is to understand a physical theory, one might be tempted to say that it amounts to mastering the mathematics involved in the formalism. According to such a rather simplistic view, physicists understand a theory whenever they know well the rules for manipulating the equations. What is required to acquire such understanding is simply to learn the math. Such an account of understanding is rather unsatisfactory. Arguably, understanding involves more than computing abilities. For example, computers compute, and often compute very well, though it is highly doubtful that they "understand" anything.

De Regt and Dieks ([33]) have given a much more interesting analysis of the kind of understanding scientists can have of a physical theory. Note that they actually give 1) a necessary and sufficient condition for a phenomena to be understood in terms of having an intelligible theory and 2) a sufficient condition for a theory to be intelligible. Our condition for T-understandability is based on what they call intelligibility of a theory:

T-understandability A theory is *T-understandable* if and only if there is a person at some time (past, present, or future) who could predict how a system in a given theoretical context is going to evolve, without any need for explicit computation[2]. Moreover, this ability has developed in virtue of the person's familiarity with the theory in question.[3]

[2]We render de Regt's and Dieks' sufficient condition necessary in adding an element of modality in the condition: they hold that it is sufficient for a theory to be intelligible that some scientist(s) is (are) able to predict without computation; we add to this that it is necessary for a theory to be intelligible that someone *could* be able to predict without computation.

[3]This last clause is not included in de Regt's and Dieks' condition for intelligibility, but is required in the case of empirically equivalent theories. One does not have T-understanding of theory A if they make predictions in accordance with A in virtue of their T-understanding of B, an empirically equivalent theory.

> Someone has actual *T-understanding* of a theory if and only if they can predict how a system in a given theoretical context is going to evolve, without any need for explicit computation and this ability has developed in virtue of the person's familiarity with the theory in question.
>
> The T-understandability of a theory for a person comes in degrees; it depends on how well they can predict how a system can evolve without the need for explicit computation.

History tells us that it takes one to two generations of physicists before a new theory to be understood in this sense. It has been the case for Newton's theory, for which, for example, the Cartesian model of contact action had to be given up in favor of action at a distance. It has been also the case for quantum theory where the twentieth century idea that physical entities are either waves or particles had to be given up. It is important to note that, if to understand quantum theory is just this, to be able to use the theory in predicting a given system's behavior without explicit computation, then physicists have certainly come to understand quantum theory by now.

T-understanding is certainly an important part of scientific activity. As de Regt and Dieks points out, it allows the physicist to understand the observable phenomena in a very specific way. Such T-understanding is sufficient for a physicist to make his way through the experiment and do most of his work. The impressive progress made during the twentieth century in quantum physics is due to the possibility and efficiency of T-understanding. The majority of quantum physicists, who are not interested in foundational issues, have proved able to work with SQM without further need for a coherent interpretation.

That said, T-understandability is not all there is in the understandability of a fully interpreted theory. The understandability of the interpretation also matters. For example, the orthodox interpretation (misnamed as it is really a separate theory form standard quantum mechanics because of the added collapse dynamics) is perfectly T-understandable, but very few physicists think it is IT-understandable. Reaching T-understanding has not been enough to reach IT-understanding. Many classes on quantum physics start with some warnings about the weirdness of the theory as interpreted in the orthodox way or even go so far to say that quantum theory is not something that should be attempted to be understood. These warnings seem inexplicable if T-understandability exhausted understandability. We also must consider I-understandability.

Understandability of an interpretation

What is it for an interpretation to be understandable? We might first ask what an interpretation does. An interpretation indicates what the world might be like if a theory is true. In doing so it describes at least some of the ontology of the world (that pertaining to the domain of the theory) that is compatible with the theory. Moreover, it also indicates how the ontology proposed can account for the macroscopic appearances, especially if it is far at odds from a classical ontology of definite-valued properties evolving in space and time.

This latter feature of interpretations is a criterion of sufficiency. It sorts potentially acceptable interpretations from unacceptable ones. Note however that acceptability and understandability of an interpretation do not necessarily go together. Of course, not all acceptable interpretations, that is, that recover the appearances, are understandable. On the other hand, it is reasonable to expect that if a person understands an interpretation, they be able to indicate how it does *or does not* recover the appearances. It is important that an interpretation is understandable even when it cannot recover the appearances, that is, when it is not acceptable. Bell's theorem provides a perfect example. Bell's theorem shows that quantum probabilities cannot be interpreted in terms of our ignorance of underlying, definite valued properties of physical systems evolving according to local, non-contextual laws. So, the understandability of an interpretation must consist in more than just saying how the appearances are recovered, or fail to be recovered by the interpretation. The understandability of an interpretation must have something to do with the understandability of the ontology associated with the interpretation as well.

We propose the following rough characterization of the understandability of an interpretation:[4]

I-understandability An interpretation is *I-understandable* if and only if there is a person at some time who could define the ontology employed by the interpretation in terms of concepts they possess and indicate how the interpretation succeeds or fails at recovering the appearances.

A person has actual *I-understanding* of an interpretation if and only if they can define the ontology employed by the interpretation in terms of concepts they possess and indicate how the interpretation succeeds or fails at recovering the appearances.

The I-understandability of a theory comes in degrees; it depends on how completely one can define the ontology in terms of concepts they possess and how completely they can indicate how the interpretation succeeds or fails at recovering the appearances.

So, understanding an interpretation consists in having the right kind of concepts with which to make sense of the world. A few examples will make the analysis plausible. In order to understand a Newtonian world, the concept of instantaneous velocity is required, as is instantaneous action at a distance, and inertia as well. To understand the world of classical electrodynamics, one requires the concept of a field. Having the right concepts goes hand in hand with the understandability of the interpretation of a theory. Then one needs to know whether or not the ontology is compatible with the appearances.

[4]Note that we do not pretend in this paper to provide a full analysis of the understandability of an interpretation. We only seek to establish rough notions of understandability to bring some clarity to the debate between Bohmians and Many-Worlders.

Our proposal here is less controversial than it might seem. As emphasized above, we do not pretend to give here a full and definitive theory of our categories of understanding. I-understanding is only supposed to characterize our understanding of an interpretation of a theory. We would not maintain that such categories of understanding, the concepts we possess to define the fundamental ontology, are fixed, quite the contrary. Also, we strictly avoid the issue of what the actual ontology of the world is.

The relationship between I and T-understandability

There is reasonable evidence to believe that I-understandability and T-understandability are independent notions, but not definitive evidence. This is because I and T understandability are threshold notions: theories achieve a sort of understandability when there exists a person that could develop certain abilities. So, simply giving examples of people who have I-understanding without T-understanding and vice versa will not be sufficient. It would have to be shown that there are people who could achieve I-understanding while no other person could achieve T-understanding of a theory, and vice versa to establish complete independence.

It is certainly plausible that I-understandability is independent of T-understandability. We only need to imagine an interpreted theory with a simple ontology which cannot possibly recover the appearances, but with a convoluted dynamics that is beyond the practical capabilities of human computational abilities. It is not clear that T-understandability and I-understandability are independent. A similar thought experiment in this case is not convincing. It's not clear that there could be a person who could develop T-understanding of an interpreted theory while it being somehow impossible for anyone anywhere at any tine to develop I-understanding of the interpreted theory.

What can be said with certainty is that *for a particular person*, I and T-understanding of a theory are independent. It is a fact that many twentieth century quantum physicists have reached T-understanding without having I-understanding. Many quantum physicists who have a T-understanding of quantum mechanics and use the theory without thinking about a coherent interpretation. Only a minority of physicists have gotten interested in foundational issues. It is striking that the majority of working physicist have actually reached T-understanding through use of the orthodox interpretation (theory), even if it is inconsistent.

I-understanding is independent of T-understanding for a particular person. Indeed, many philosophers know the details of the possible ontologies corresponding to the different interpretations and can indicate how the ontology can succeed or fail in accounting for the appearances. That said, few would pretend to be able to use the theory as professional physicists do, i.e. given a system in a certain theoretical context, to be able to predict how the system is going to evolve without explicit computation. This seeming independence of I and T understandability allows us to make sense of claims regarding relative understandability made by Bohmians and Many-Worlders.

2.2 The Bohm vs. Many-Worlds debate over understandability

In this section we argue that BIT is more easily I-understandable than MWIT. We also argue that MWIT is more easily T-understandable than BIT. We will argue that these facts do not allow us to say definitively whether BIT or MWIT is more IT-understandable. Moreover, these facts regarding I and T-understandability alone do not allow one to decide the debate.

2.2.1 Bohm vs. Many-Worlds: T-understandability

MWIT contains standard quantum mechanics. It does not add anything to the standard formalism and basic correspondence rules. Since SQM is T-understandable, MWIT is T-understandable. Now, BIT is not restricted to the standard formalism. It contains the guiding equation in addition. Though the introduction of the guiding equation does not necessarily render BIT unT-understandable, it certainly makes it more difficult for people to develop a T-understanding of the theory. Explicit calculations are typically more difficult with Bohm's theory. It is no great leap of faith that this hinders the T-understandability of the theory. So, the MWIT is more easily T-understandable than BIT.

2.2.2 Bohm vs. Many-Worlds: I-understandability

Bohmians are well known to claim that their interpretation is best because it has a "clear ontology": particles with definite trajectories in space and time. A typical example of such a claim is found in the abstract of a famous paper by Dürr, Goldstein, and Zanghi:

> We argue that [the quantum formalism] can be regarded, and best be understood, as arising from Bohmian mechanics, which is what emerges from Shrödinger's equation from a system of particles when we merely insist that "particles" means particles. While distinctly non-Newtonian, Bohmian mechanics is a fully deterministic theory of particles in motion, a motion choreographed by the wave function. [...] When the quantum formalism is regarded as arising in this way, the paradoxes and perplexities so often associated with (nonrelativistic) quantum theory simply evaporate. ([40, abstract])

In what sense can an ontology of particles moving in space and time be said to be more easily understandable than an ontology of worlds with internal observers? In the sense of I-understanding as defined above. The concepts in which the ontology that BI employs is definable are widely possessed. The BI employs an ontology consisting of particles and a field which acts non-locally. We all know what particles and fields are. The fact that the field is non-local presents no special conceptual challenge as we are perfectly familiar with non-locality from Newton's theory. Moreover, probabilities present no special problem for the BIT. The theory is deterministic and probabilities receive an ignorance interpretation, just as they do in statistical mechanics. Indeed, BI is easily I-understandable.[5]

The strength of the BIT here is the MWIT's weakness. The appearances are Newtonian. The world appears to consist in definite valued properties that evolve deterministically in space and time. The difficulty for Many-Worlders is that they give up definite values for properties. In taking the wave function as a complete representation of the world, the Many-Worlders accept superpositions, and hence indefinite properties – spin and position for instance, in their ontology, and this both at the fundamental and phenomenological levels. Moreover, they have to account for probabilities, but cannot do so in terms of an ignorance account because generally all values associated with the measurement of a property are realized in some sense, but not in the sense of possessing sharp values. The Many-Worlder faces the task of reconstructing the appearances with tools that could hardly seem more unsuitable.

That the ontology proposed seems so at odds with the appearances entails nothing about the understandability of the interpretation. The problem lies in the fact that no one seems to be able (for now) to define the ontology in terms of concepts that are possessed widely, or even at all. Many-Worlders will often claim that the only ontological features of the world are those that correspond to the wave function, but that existential claim does nothing to define what the ontology is. One can say that the world is composed of non-definite properties,

[5]One might point out that contextuality is an interpretational oddity of Bohm's theory as an objection to the claim that it is I-understandable. While it may be an ontological oddity in relation to what we come to expect from the appearances, contextuality itself is perfectly commonplace. E.g. The probability that a certain horse will win a race will depend on the other horses in the race. So, in terms of the analysis of I-understandability offered above, it is not a challenge for the BIT.

but simply because we possess the concepts in which "definite property" can be defined does not mean that those same concepts can be used to define non-definite properties. It is this fact that challenges the I-understandability of the MWIT. The fact that arguably no one I-understands the MWIT is, of course, no guarantee that it is not I-understandable. That said, on this matter, clearly the BIT is more easily I-understandable than the MWIT.

2.2.3 Bohm vs. Many-Worlds: minimality

No argument in favor of MWIT can be made on the basis that the MWIT is easily I-understandable. However, one often encounters the argument that the *minimality* of the MWI *vis à vis* the formalism makes it preferable to Bohm's theory. It will be argued that minimality has nothing to do with the I-understandability of an interpretation.

The theory, because it works wonderfully as far as empirical adequacy is concerned, is good enough for what physicists do. Further, they know by now so well how to use it, that they do their job without being puzzled by the quantum behavior anymore. Given such T-understandability of the theory, it is often argued that the most easily understandable interpretation would be one that takes the formalism at face value; namely, that which is *minimal* vis à vis *the formalism*. An interpretation is *minimal* vis à vis *the formalism* only if it only postulates as fundamental entities the counterparts of some abstract mathematical objects inhabiting the formalism (there always may be mathematical artifacts in the formalism). The Many-Worlds interpretation is one such interpretation. It does not postulate any extra elements in the ontology besides the ontological correspondent of the universal wave function. By contrast, the argument goes, Bohmians postulate particles with definite trajectories, hence at least one non-reducible definite valued property, which is not needed for an empirically adequate theory. Instead, Many-Worlds interpretations thus only juxtapose the simplest ontology there is to the formalism. To adopt a Bohmian-type interpretation thus involves an inflated ontology.

Minimality may be a virtue of an interpretation. That said, it is clearly a separate virtue from I-understandability as defined above. The MWI is not easily I-understandable, despite it's minimality. Minimality seems to respect a principle of ontological parsimony that is a guiding rule in metaphysics. Postulate no more than needed to explain what needs to be explained (the appearances). However, whether an ontology is minimal or not has nothing to do with whether it is easily understandable. A world in which every physical quantity has a definite value is less minimal than a world in which many physical quantities have contextually defined values that are reducible to other physical quantities, as in Bohm's theory, despite the fact that the latter is less easily understandable. So, minimality has nothing to do with the debate over the understandability of the MWIT or BIT.

2.2.4 Bohm v. Many-Worlds: IT-understandability

What we have argued so far is the the MWIT is more easily T-understandable than BIT, but BIT is more easily I-understandable than MWIT. Claims regarding the understandability of MWIT and BIT are not that specific, but general. What we can tell regarding the comparative IT-understandability of these interpreted theories from those facts? What weighting does one assign to I and T-understandability to determine IT-understandability? We will argue that T-understandability is more important than I-understandability, but even so, it doesn't give us a reason to prefer BIT over MWIT.

The T-understandability of a theory seems quite important. If physicists cannot T-understand a theory, it is likely that their creative abilities and insights will be hampered from always having to do explicit calculations. Moreover, as an eminently practical matter, the more easily T-understandable a theory is for students, the better, as it will quite literally give them more time for research.

The I-understandability of a theory seems less important. To frame the issue, one might ask, "What can a physicist *do* with I-understanding that can't be done with T-understanding?" After all, the physicist with T-understanding of a theory will be able to employ a theory to empirical effect in a way a physicist with only an I-understanding of a theory could not. If the aim of physics is to effectively interact with the world, it would seem that T-understanding is more important than I-understanding.

One thing that can be said of I-understandability is that it is helpful for developing a T-understanding of an interpreted theory. When one has a sense of what the abstract formalism of a theory corresponds to, one has an easier time learning the formalism (There is a reason why we learn arithmetic by thinking about apples and oranges.). Moreover, when one has a representation of what happens in the world during certain kinds of physical interactions, perhaps a visual representation, one can more easily acquire the ability to make predictions without explicit calculations.

Note that we are not suggesting that I-understandability isn't important. Obviously, reflections on the I-understandability of interpreted theories can lead to new research programs and even new theories. Perhaps Bohm was dissatisfied with the understandability of the Copenhagen interpretation and this motivated him to develop his theory of quantum phenomena.

At this juncture, philosophers might want to invoke Einstein's reflections on the conceptual foundations of electrodynamics and it's importance for the development to the theories of relativity, to emphasize the importance of I-understandability. It is important to keep in mind that the understandability of an interpretation is not correlated to the acceptability of an interpretation, e.g. modal interpretations of quantum mechanics (which have severe problems but are I-understandable). Einstein was not satisfied with the interpretation of electrodynamics, and finding the flaws in that theory spurred great breakthroughs. That said, electrodynamics was perfectly I-understandable and understood by a great many physicists.

To sum up, T-understandability is more important than I-understandability and should contribute more to IT-understandability than I-understandability does. One might think that this decides the BIT v. MWIT debate in favor of MWIT. The problem here is that the difference in T-understandability is slight, and the difference in I-understandability is more substantial. So, even if T-understandability is weighted more heavily than I-understandability, the IT-understandability of MWIT and BIT seem to be roughly equal. Understandability does not appear to give us a means of preferring one interpreted theory over another in this case.

2.3 Understandability vs. Truth

The above analysis of I-, T-, and IT-understandability of physical theories provides a reasonable framework for assessing and comparing the understandability of theories. Even though the determination of the IT-understandability of BIT and MWIT failed to be decisively in favor of one interpreted theory or another, it would appear that the IT-understandability could be one of many criteria of choice for interpreted theories along with accuracy, simplicity, scope, etc. In this section, we will actually suggest that understandability of a theory or of an interpretation is a weak criterion of choice, especially if our aim is truth.

In formulating physical theories, we minimally aim at true observable predictions. A physical theory must save the phenomena of its domain. T-understandability of the physical theory does not weigh very much in comparison to empirical adequacy. For example, a theory that implies false predictions cannot be accepted on the argument that it is perfectly T-understandable. Conversely, were we given, say by some god, the true theory of the world, that is, as we defined it, a formalism with the basic correspondence rules that allows us to apply the formalism to make empirical predictions, there is little doubt that we would accept and use it, even if, due to our cognitive insufficiencies, we were doomed to never T-understand it and to go through painstaking computations forever. In such a case we might also want to utilize a more easily T-understandable theory for practical matters, much as we use Newton's theory for many calculations where relativistic effects are negligible. The point is that we would certainly never reject the true theory of the world in virtue of it's T-understandability.

Many would further maintain that, in designing physical theories, we ultimately aim at finding the true laws that are behind the observed regularities in the world, not only finding empirically adequate laws. Of course, we can never be assured that we possess the true laws in hand because several sets of laws are compatible with a set of phenomena. That said, it does not follow from this that the true laws are not what we are aiming at as an ideal. If true, whether the dynamical laws a theory contains be T-understandable or not should not weigh much when it comes to accepting it.

The same argument holds in the case of interpretations. In formulating interpretations, we hope to grasp something of the fundamental ontology of the world. Granted again, we do not have any empirical access to such ontology. Neither do we have the means to test an interpretation besides its consistency with the observable predictions of the theory. This does not imply that we are not trying to get as close to the true story of the world as we can. For example again, we cannot see anyone adopting an interpretation which we would know is false, on the basis that it is highly understandable. Conversely, were we given, by some goddess this time, the true ontology of the world, there is little doubt that we would use it, be it in terms of definite-valued properties distributed in space and time or not. Moreover, such knowledge would likely contribute to the development of our theories.

In the case of theories as in the case of interpretations, understandability thus is neither necessary nor sufficient as a criterion of choice. This is because theories and interpretations are, ideally, designed to be true of the world, and that understandability is neither necessary nor sufficient to warrant anything like truth. That it is not sufficient is clear enough. All our false but understandable theories are obvious examples. It is not necessary either, for it could be well the case that the fundamental laws, if any, and the fundamental ontology cannot fit into the framework we happen to use to understand the interactions between physical systems at the observable level.

The understandability of an interpreted theory obviously depends on our cognitive abilities, as twenty-first century adult human beings. Such cognitive abilities are highly contingent. They arguably depend on the average scale of the objects with which we usually interact, and on the evolution of our sense apparatus and neural connections, which make us interact in a particular ways with such objects. The relative understandability of theories and interpretations could be ranked completely differently by an organism which was equipped differently than we are, either physically or experientially. If true, then it seems besides the point to consider understandability as a criterion of choice for theories and interpretations. In short, the type of environment in which we happened to evolve does not have much to do with the truth of the matter about what the world is ultimately like.

2.4 Conclusion

We have tried to clarify the debate over the respective understandability of the Many-Worlds interpretation of the standard formalism and Bohm's theory. We have understood claims on each side of the debate regarding superior understandability as reconcilable on the notions of I and T-understandability developed in this paper. BIT is more easily I-understandable, and the MWIT is more easily T-understandable. In terms of IT-understandability, the MWIT and BIT seem to be on roughly equal footing. So, we cannot advocate one of these two interpreted theories over the other on the basis of its higher degree of understandability. Moreover, even if there was a great difference in IT-understandability between them, it would be a weak reason to prefer one over the other, as IT-understandability is independent of truth, the ultimate aim in accepting theories for most of us.

Chapter 3

Understanding made modest: scientific theories as families of models

3.1 Introduction – Semantic vs. Syntactic

The semantic view was born historically as well as conceptually through its opposition to the syntactic view. The syntactic view of scientific theories has been labeled the "received view", because it has been the most important trend in the philosophy of science for more than half of the last century.[1] According to the syntactic view, a scientific theory is defined as an axiomatic system of formal sentences, which are formulated in theoretical language, and partially interpreted in terms of observational language by means of rules of correspondence. The syntactic view is a normative, formal and linguistic-oriented account of scientific theories. Only theories which can be given such axiomatic formulation are considered scientific. Lastly, the syntactic view focuses on the products of scientific activity, scientific theories, which are taken to exhaust all there is in scientific knowledge that can be interesting from a philosophical point of view. The proponents of the semantic view consider the syntactic

[1]See [91], [24], [65] and, for an historical presentation of syntactic view, [109].

program to be a failure for two reasons.[2] One is that the defenders of the program do not seem to have found any satisfactory formal characterizations of scientific theories. The possibility of formulating a theory in the prescribed canonical way is neither necessary nor sufficient a condition to qualify the theory as scientific.[3] Another reason is that in looking for such a characterization, they got tangled in endless discussions of difficulties imported from metalinguistics, which are irrelevant to philosophy of science for they stand "à mille milles de toute habitation scientifique."[4] Thus, the criticism is not only that the syntactic research program does not achieve any satisfactory conclusion. That alone would not be enough for giving up a research program in philosophy. Rather, the criticism is that these endless controversies, which stem from a linguistic orientation, are irrelevant; they fall outside the proper domain of philosophy of science, that is, scientific theories as they are constructed, used and tested in actual practice.

The semantic view has two main features in response to the two criticisms above. First, the semantic view distinguishes scientific theories from particular (linguistic) formulations of theories. Theories are to be studied from an extra-linguistic point of view. Second, the semantic view aims at a descriptive account of scientific theories as opposed to a normative account as the received view attempts to give. In both setting aside matters concerning specific formulations and normative characterizations of scientific theories, the semantic view is intended to give an account of scientific theories more adequate to scientific activity.

The turn from the syntactic to the semantic view thus crucially consists in shifting from a normative theory of the specific linguistic formulation of scientific theories to a descriptive analysis of scientific theories as used in actual practice in non-linguistic terms. The central claim is that focusing on models of theories is the best way to give such analysis. Within the semantic view, a theory is not defined as an axiomatic system of sentences, but as the

[2]See [112], [56], [128], [110].
[3]See [79].
[4]From [130, 225], quoting (in French) Saint Exupéry, [34].

family of models of its possible formulations. The notion of model appealed to is the logical notion of a model, typically referred to as the Tarskian notion of model.[5]. A model is the set-theoretical structure that gives an interpretation to the possible sets of sentences in which a theory is formulated so that the sentences are true on the interpretation. Thus, the semantic view of scientific theories was called "semantic" for the obvious reason that it attends to the semantics of theories.

The focus on models is supposed to address the two criticisms raised above. First, focusing directly on models allows one to get rid of the problems imported from metalinguistics that plagued philosophy of science during the syntactic reign. Of course, it should be noted that if the notion of model taken is a linguistic notion (such as definitions, descriptions or the like), the semantic view is an obvious target for the criticisms marshaled against the syntactic view. The focus on models is, on the contrary, aimed to allow for a non-linguistic account of the theories. Further, models are always interpreted. The question of interpretation, which stemmed from the distinction between theoretical and observational languages in the syntactic view, just vanishes. A few points of clarification are needed. The point of focusing on models is neither to avoid formulations of theories altogether, nor to deny that to put a theory in axiomatic form can be philosophically interesting. It might be noted that, formally, to consider a theory from the point of view of the family of its models does not make much difference from considering it from the point of view of the set of its axioms. Proponents of the semantic view certainly agree with this, but this is not the point. The point is that no specific formulation of a theory is privileged. The change is a change of perspective, nothing more, but nothing less.

Second, the semantic turn is supposed to give philosophy of science the opportunity to re-focus on its proper domain. A central claim of the semantic view is that to consider theories as families of logical models is a way to analyze scientific theories that is adequate

[5]See [82], [120], and see [55] for a rigorous presentation

to actual practice. Indeed, "models" are what scientists use when learning, constructing, testing and developing scientific theories. That said, the central claim above amounts to claiming that the models used in scientific practice are logical models in some sense. This thesis, which I shall call the identification thesis, has been the center of the controversies and criticisms of the semantic view for the past ten years.[6] Many have argued that the two notions of model are irreconcilable. Some have concluded that the semantic view as a research program is flawed; others have tried to make sense of the program by focusing only on the notion of scientific models. One aim of this paper is to take identification thesis seriously, to try to understand its motivations within the semantic view, and to assess in what sense and to what extent it is central to the view. Such analysis should allow us to show that most criticisms stem from a misinterpretation of the semantic view, that is an interpretation which is arguably unfaithful to its initial project, as defined by Suppes in the sixties and as developed by him and others since. Let me distinguish between two possible interpretations of the identification thesis. A strong interpretation of the identification thesis is:

When scientists develop and test a high level theory, the models that they construct and use are the logical models of the high level theory, and as such model theory is a fruitful tool for philosophers to give an account of the actual practice of science.

This strong interpretation seems to be at stake in the debates over the semantic view and will be the subject of the next section. However, I shall defend later a more modest interpretation of the identification thesis, according to which what the semantic view really claims is this:

When scientists develop and test a high level theory, the models that they construct and use are analyzable in terms of the logical models of a hierarchy of theories, and as such model theory is a fruitful tool for philosophers in the formal analysis of scientific theories.

[6] See [26], [86], [108], and [125].

The shift from the strong to the modest interpretation of the identification thesis is twofold. First, there is a weakening of the thesis insofar as all the models that scientists use are not viewed as the models of the high level theory. Second, it is a shift to a methodological point of view. I shall defend the idea that the semantic view is best seen as the methodological prescription to use model theory for a formal analysis of scientific theories.

In Section 3.2 I 1. give an argument for why some form of the identification thesis is central to the semantic view, 2. briefly expose the criticisms that were recently raised against the tenability of the identification thesis and, consequently, 3. state the dilemma that the semantic view must address if these criticisms are successful. In Section 3.3, I argue that the criticisms leveled at the identification thesis only apply to the strong form of the thesis. Hence, under the modest interpretation of the identification thesis, the dilemma is bypassed. In Section 3.4 I briefly discuss the most controversial consequence of the modest interpretation of the identification thesis; namely, that the semantic view does not address the problem of representation. I argue that the semantic view does not have to address this problem, but that it still provides a useful characterization of scientific theories.

3.2 From the identification thesis to the dilemma

The identification thesis stems from two requirements that are constitutive of the semantic view. First is a requirement of adequacy. The semantic view aims at giving an account of scientific theories as they are used in actual practice. As presented above, the requirement of adequacy is a central motivation of the semantic view. Proponents of the view claim that it is a fact of the matter that most scientific activity consists in the consideration of models, rather than the axioms of a theory. To fulfill the requirement of adequacy, it is thus essential to the view to consider scientific models, that is, models that are constructed, used and tested in scientific practice.

The second requirement is that scientific theories should be given a systematic, formal analysis. Let us call it the requirement of formalization. The semantic view shares with the syntactic view the aim of giving a formal analysis of scientific theories. The semantic turn prescribes a change in tool, not in the kind of analysis. Granted, the analysis is not to be normative, but this is another point. The point here is that the proponents of the semantic program still believe in the necessity of a formal framework for the systematic treatment of issues in philosophy of science and scientific methodology. However, first order logic plus identity, advocated as the tool to use to analyze theories on the syntactic view, is just too poor a tool to give an account of the complexities of scientific theories. The upshot is that the syntactic view could not consider anything but oversimplified theories. Adding the axioms of (a fairly uncontroversial version of) set theory is just what is needed to be able to give an account of most important scientific results, because it includes most of mathematics. Proponents of the semantic view advocate model theory as the correct framework for a formal analysis of scientific theories. Consequently, it is essential to the view that the models considered are models in the logical sense.

The requirement of adequacy and the requirement of formalization together thus make the identification thesis central to the view. Recent literature has challenged the identification thesis under its strong interpretation. Critics aim to show that models used in scientific practice are not models of the theory, in the following sense: scientific models, which are used as representations of phenomena, do not have some properties which are characteristic of logical models and vice versa. Logical models are essentially characterized by their truth-making function. Several have argued that scientific models used for the domain of some theory are not the truth-makers of the theory. Redhead provides an analysis of the relationships between models and theory.[7] One important class of models Redhead identifies as "impoverishment models". What characterizes impoverishment models is that they

[7]See [92].

contradict the theory. They arise from the simplification of the equations of the theory considered in order to get the solutions more tractable. Such models are disconnected from the theory in so far as the theory does not provide the means to mathematically justify the approximations. Moreover, the simplified equations generally have quite different solutions than the original equations. The point can be reformulated by saying that some models feature real distortions with regards to the theory that are not reducible to limit cases or Aristotelian abstractions.[8] Accordingly, such models simply cannot be true of the theory.

If scientific models do not function as logical models do, then how are they to be understood? According to Morrison, models are "autonomous agents in the production of scientific knowledge."[9] Morrison studies the "production of scientific knowledge" from two points of view: the construction of and the function of models in practice. In both cases the main claim is that the way scientific models work is not reducible to them being models of the theory. Concerning the construction of models, Morrison's main thesis is that models cannot be derived from a theory through any systematic process. If scientific models were models of the theory, according to Morrison, the theory should provide sufficient means to construct these models. On the contrary, argues Morrison, actual construction of models is an "art" that demands something else than formal analysis of the theory. Concerning the way models function in practice, Morrison's main thesis is that they produce supplementary knowledge that the "abstract theory" cannot provide. They function as "mediators" between experiment and theory, making explicit certain "structural dependencies". Models specify what adjustments and modifications are needed to get the experimental results to fit the theory. And this is simply the kind of knowledge that the high-level theory does not provide.

From a more conceptual point of view, Thomson-Jones proposes the following analysis. He distinguishes between the two notions of models according to their functions, as truth

[8]See [27] for an exposition of the distinction between (Galilean) idealization and (Aristotelian) abstraction.
[9]See [87, 39].

maker (logical model) and as a representational structure (scientific models).[10] He claims that the essential function of scientific models is to represent. Further, he demonstrates that the logical function of models is not enough to give an account of the representational function of scientific models. Thus the semantic view, if it sticks to the identification thesis, cannot account for the essential representative function of scientific models.[11]

According to the analyses above, the two notions of models are seemingly irreconcilable. On the one hand, scientific models do not seem to be necessarily truth-makers of the theory. On the other hand, the truth-making function of logical models does not seem to be sufficient to give an account of how scientific models are constructed and used in scientific practice. If true, the semantic view faces the following dilemma: either it fulfills the requirement of formalization, i.e. remains faithful to the logical notion of model, but then cannot claim to fulfill the requirement of adequacy, or it sticks to the requirement of adequacy, but then the model-theoretic notion of model is unacceptable, and the semantic view has to give up being 'semantic'.

One escape is to give up on the 'semantic' aspect, that is, to give up the logical notion of models and focus on the representational aspect of scientific models.[12] I have argued that this would amount to giving up the semantic program altogether, for the consideration of models in the logical sense stems from the requirement of formalization, which is central to the view. In the next section I argue for an alternative escape from the dilemma. I relax the strong interpretation of the identification thesis in favor of a more methodological stance.

3.3 Methodological semantic view (MSV)

Let us reconsider the distinction made at the beginning of the paper[13]:

[10]See [125].
[11]See [20] for another clear exposition of this point.
[12]Thomson-Jones does this in [125].
[13](S) stands for 'strong' and (M) for 'modest' or 'methodological'.

(S)

When scientists develop and test a high level theory, the models that they construct and use are the logical models of the high level theory, and as such model theory is a fruitful tool for philosophers to give an account of the actual practice of science.

(M)

When scientists develop and test a high level theory, the models that they construct and use are analyzable in terms of the logical models of a hierarchy of theories, and as such model theory is a fruitful tool for philosophers in the formal analysis of scientific theories.

Section 2 has made it clear that (S) is not tenable. I shall argue not only that (M) is tenable, but characterizes a research program for philosophy of science which is worth considering. I shall call this the methodological semantic view (MSV). The basic objection against the possibility of taking a semantic perspective on scientific models is that these scientific models do not feature the characteristic properties of logical models. In particular, they are not models of the theory in the sense that they are not true of an abstract, general and unified theory. A naive picture of how theories relate to the world would be the following. A high level theory, such as quantum chromodynamics, is characterized by its models. Given some phenomena in the world, such as the results of scattering experiments, the theory has a model that represents the phenomena of interest. The authors considered in the last section have shown that this view is simply incorrect. Models that scientists use to represent the phenomena are not the (logical) models of the high level theory.

A more realistic picture is that a whole hierarchy of models stands between the phenomena and the high level theory. In virtue of the definition of the model, there is a theory on which the model is true at every level of the hierarchy. Note that this is not claiming that we have such theories at hand, but rather that a model-theoretic perspective can be taken on the various scientific models. The semantic view, as defined originally, has never been not committed to the claim that all models in this hierarchy are logical models of the "high level" theory. Suppes made this point clear already:

> What I have attempted to argue is that a whole hierarchy of models stands between the models of the basic theory and the complete experimental experience. Moreover, for each level of the hierarchy, there is a theory in its own right. Theory at one level is given empirical meaning by making formal connections with theory at a lower level. Statistical and logical investigation of the relations between theories at these different levels can proceed in a purely formal, set-theoretical manner. The more explicit the analysis the less place there is for non formal consideration. Once the empirical data are put in canonical form (at the level of models of data) every question of systematic evaluation is a formal one.[14][15]

So, the criticisms of the semantic view considered in the last section would only apply to the semantic view if it included the strong form of the identification thesis. The appropriate characterization of the semantic view is methodological. Properly understood, the semantic view consists in considering directly the whole hierarchy of scientific models and use model theory for a formal analysis of these models and of their relationships. Seen as the methodological prescription to use model theory for a formal analysis of the hierarchy of models that scientists consider, the semantic view bypasses the dilemma and is fully tenable. Now, what is it good for? Although much of the debate in philosophy of science recently focused on the question of representation, I shall claim that MSV consistently holds that, leaving aside this question, there is room for a formal analysis of the structure of theories in terms of the relationships between their models. The point of MSV is to claim that taking the perspective of model theory can give insights on scientific models, leaving aside their representational function. I turn to this in the following section.

[14] [111, 34].

[15] Also see [20].

3.4 Cleaning up: what issues MSV addresses or not.

Philosophy of science is essentially concerned with inter-theoretic, intra-theoretic and theory-world relations. MSV can fruitfully address the first two issues, but not the third one, which includes the representational function of theories or models. If the semantic view is a methodological prescription for using model theory as a tool for the formal analysis of scientific theories, then it stays in the semantic level. MSV studies the relationships between models, from the more abstract to the models of data, but never can nor ever has to go beyond the models, and hence, beyond the formal existence of the objects the theory presents.[16]

One might ask in what sense the semantic view is useful if it does not analyze theory-world relations. No one has said it better than Suppes: "the notion of model in the sense of logicians provides the appropriate intellectual tool for making the analysis both precise and clear."[17] In particular, it makes the notion of structure and of the structural relationships between models precise. Such a precise and well-defined tool is useful in addressing both issues of internal structure of theories and interrelations between theories. For example, van Fraassen in [131] is an explicit application of the method presented in his [130], to examine the controversies that have plagued the philosophy of quantum mechanics since the inception of the theory. Suppes shows how a proving representation theorem with model theory is useful to address the question of reductionism.[18] Indeed, to prove a representation theorem is to show that there is a specific class of models such that every model of the theory is isomorphic to a member of this class. This is arguably an interesting way to formalize the possible reduction of the theory to this restricted class of models. Lastly, using the resources

[16]For more details on the distinction between presentation and representation, in [20]. For an exposition of how the semantic view consider scientific theories as deploying different levels of models, see [79].
[17][113, 17].
[18][116].

of model theory has been the only means to make structural realism a precise view.[19] Note that an advantage of MSV is that, staying at the semantic level, it does not entail any position in the realist/antirealist debate. Thus, MSV can be useful on for both sides of the debate.

There is no question that scientific models have a representational function. If the semantic view is supposed to be a comprehensive view of science, then it had better say something about representation. However, the semantic view was never intended to give a comprehensive view of science. Consider the following quote from Suppes:

> It is a myth engendered by philosophers - even in the past to some extend by myself - that the deductive organization of physics in nice set theoretical form is an achievable goal. A look at the chaos in the current literature in any part of physics is enough to quickly dispel that illusion, this does not mean that set theoretical work cannot be done, it is just that its severe limitations must be recognized.[20]

According to me, the severe limitations that have to be recognized are just that the semantic view cannot address the question of representation. That said, just as Suppes remarks: this does not mean that no set theoretical work can be done.

We began this paper by considering a crucial objection to the semantic view. If the semantic view involved the claim that the entire hierarchy of models used in scientific practice are models of a supposedly unified high level theory, then it would be untenable. As remarked above, the semantic view of theories is not committed to that thesis. I argued that only a modest version of the identification thesis is central to the semantic view. The semantic view is best seen as a methodological prescription to use model theory for a formal analysis of the hierarchy of models that scientists consider. MSV is both tenable and still a promising

[19] See [32].
[20] [115, 214].

research program for philosophy of science. This legitimates the use of model theory in philosophy of science.

Part II

Understanding Bell-type phenomena

Chapter 4

The mainstream interpretation in trouble?

Bell-type phenomena have been the subject of much philosophical investigation for the last twenty five years. It is commonly held that the discussion has reached a satisfactory conclusion. It is widely believed that Bell-type experiments force us to accept that the quantum domain cannot be described by any theory that gives a representation of the phenomena as local causal processes. That said, according to the mainstream interpretation, the form of non-locality that Bell-type phenomena force us to accept is benign in the sense that it does not necessarily conflict with the locality constraints that many take relativistic theories to impose on causal processes. Put another way, it is generally agreed that our two favorite physical theories stand in a state of "peaceful coexistence".[1] Despite this apparent consensus, criticisms have been recurrently leveled against the mainstream interpretation in the literature. Our aim in this part of this dissertation is to assess rigorously whether the mainstream interpretation is well founded or not. This chapter serves as an introduction to our analysis of the interpretations of Bell-type phenomena.

[1] The phrase is Shimony's.

We shall begin with giving some background information on Bell-type theorems and Bell-type phenomena (Section 4.1). With this in hand, we shall present the mainstream interpretation, its usual justification and the ongoing, albeit underground criticism leveled against it (Section 4.2). In Section 4.3, we explain how some of these criticisms succeed in showing that the justification usually given in favor of the mainstream interpretation fails to support it. That said, this does not prove that the mainstream interpretation is false either, just that it remains ill-founded. In the last section of this chapter (Section 4.4), we present our program for assessing the mainstream interpretation. Such a program stems from the observation that, in the Bell literature, locality and causality are mostly defined solely in terms of restrictions on conditional probability distributions over events. Arguably though, whether or not locality and causality can be defined in terms of probability conditions is not trivial. Our general strategy is thus to consider rigorous theories of locality and of causality in order to investigate whether the mainstream interpretation holds under any of these theories.

4.1 Bell-type theorems and Bell-type phenomena

4.1.1 Introduction

Bell-type theorems come in various versions, but most of them follow the general scheme:[2]

- (I) Exposition of a set of assumptions S;
- (II) Derivation of some inequalities (the Bell Inequalities – henceforth BI) from this set;

[2] The literature on Bell theorems is plentiful. Bell's papers are indispensable references: [12],[11],[15],[14]. All of these are reprinted in the collection [14], with some others in the second edition [16]. For a recent synthesis, see [104].

- (III) Exhibition of quantum statistical predictions that violate the BI.[3]

The conclusion of such theorems follows by a simple modus tollens: not all statistical predictions of quantum mechanics are compatible with the S-type models, i.e., models which satisfy all the assumptions in S. To say it the other way around: no model of S-type can recover all statistical predictions of quantum mechanics.

It so happens that the BI are testable. Experiments which arguably correspond to Bell-type situations have been carried out many times since the early eighties.[4] These experiments are almost universally taken to favor quantum mechanics.[5] That is to say, it is widely agreed that we have strong empirical evidence that the BI are violated by quantum phenomena. The upshot is thus that no model of the S-type can give an account of all quantum phenomena.

The interpretation of the experimental violation of Bell-type inequalities varies depending on the version of Bell-type theorem considered. Bell-type theorems mostly vary with the set of assumptions considered in the derivation. Most of these sets however are supposed to capture the idea of "Local Realism". "Local realism" is a widely used, even if largely misleading, phrase for the conjunction of two conditions[6]:

Definition 1 – ***Local Realism***

A probabilistic model for Bell-type situations is a local realist model if and only if it satisfies the two following requirements:

1. *Realist: measurement outcomes are determined before the actual measurement process;*

[3]Some versions deviate from this general scheme: see [104], section 6, and references therein.

[4]The experiment performed in 1982 by Alain Aspect and his team in Orsay is considered the classic ([8], or, for a more popular presentation [7]). Redhead provides a table of the experimental results concerning the violations of the Bell Inequalities up to the last eighties in [93, p.108]. Many others have been conducted since, among which the ones by Zeilinger's team in Vienna and Gisin's team in Geneva are well known (see for example [134], and for a less technical presentation, [141]).

[5]Some loopholes, such as the detection loophole, have still not been completely ruled out. We shall not consider these potential problems for the implementation of the experiments in this dissertation in any detail. That said, such loopholes relate to Fine's Prism Models which we shall discuss in Chapter 5.

[6]See for example [73, p.61-2] for a "deliberately vague" characterization of local realism.

2. *Local: measurement outcomes are locally determined; namely, the results of a measurement are independent of what is happening on remote systems.*

In the following, we shall try to make these two conditions more clear.

4.1.2 Determinateness, Determinism and Value Definiteness

It is well known that the quantum state, given as in standard quantum mechanics, does not allow one to predict definite outcomes for a given experiment: it only gives a probability distribution over a spectrum of possible outcomes. The orthodox interpretation takes it that no more complete specification of the physical situation can be made. In other words, according to the orthodox interpretation, quantum probabilities are not merely epistemic: there are irreducible.

By contrast, hv theories take it that the measurement outcomes are determined before the actual measurement process – this corresponds to the "realist" part of Local Realism as defined above. They supplement the quantum state of the system with extra variables, which are called "hidden" variables because we do not have direct access to their values, nor do we have control of them in practice. The completed states are usually denoted as λ.[7] The complete state of the system, possibly together with the experimental context, determine the outcome of a given experiment. That said, the sense in which the outcomes are determined comes in various ways. One distinguishes between :

1. determinism,
2. determinateness,
3. value definiteness.

[7] Whether λ corresponds only to the supplementary variables, or to the completed state of the system, i.e. the quantum state plus the supplementary variables, varies in the literature. We shall take the latter option in the following.

Let us make the distinction clear by way of the following definitions.

Definition 2 – ***Determinism***

A physical transition between the times t and $t + \delta t$ is said to be deterministic if and only if the state of a closed system submitted to this transition is completely determined at a time $t + \delta t$, given its state at t.

Definition 3 – ***Determinateness***

Determinateness is satisfied if and only if the measurement outcomes are determinate, namely each experiment will have definite results, which may be fixed deterministically or only stochastically by the state of the system (and of the measurement context) before the measurement.

Definition 4 – ***Value Definiteness***

Value definiteness is the requirement that the state of a system assigns definite values for all quantum observables, irrespective of whether a measurement is performed.

Value definiteness, determinism and determinateness do not have the same status in the derivation of BI. As for determinism, it is not always part of the set of assumptions S used to derive the BI. Whereas it was assumed in the original proof of Bell's Theorem[8], the result was further generalized to the stochastic case by Bell himself[9] and by others[10] Determinism entails determinateness, but the converse does not hold in general. Note that Fine shows that any stochastic model satisfying the BI is equivalent to some deterministic hidden variable model. This mathematical result does not impair the philosophical significance of the generalizations of Bell-type theorem to the stochastic case, as Butterfield explains.[11] We shall describe

[8] [12].
[9] [11].
[10] See [29],[28], [6], and [85].
[11] [p. 77-8][22].

stochastic models in more detail in this chapter, Subsection 4.1.5. We shall give our own analysis of Fine's result in Chapter 5.

Value Definiteness is also stronger than determinateness, at least if we assume that measurement results are themselves encoded in the values of some observable. It is ruled out by the impossibility theorems of non-contextual hv models.[12] We shall not present these theorems in this dissertation.[13] That said, more has to be said about the various notions of contextuality in the Bell literature. This is the burden of the next subsection, i.e. Subsection 4.1.3. Before, let us conclude from the presentation above that the sense in which hv theories generally take the measurement outcomes to be determined is the sense of determinateness. They aim at supplementing the quantum state with these values that are missing in order to recover a determinate-outcome model of the phenomena.

4.1.3 Contextualities

Contextuality is a common issue in the Bell literature. Generally speaking, the idea of contextuality is simply that the outcomes of a given experiment depend on the specific characteristics of the experiment. Formulated like this, it seems both reasonable and simple. Whether you can win a marathon depends on who is running the race (the other quantities measured) and what the specific conditions of the race are (the weather or the humidity of the track for example). That said, there are various notions of contextuality in the literature, and it is not always clear how these notions relate to each other. The following is an attempt to draw a conceptual map of these various notions of contextuality.

The most common notion of contextuality for quantum probabilities and hv theories is the following. A hv theory is said to be contextual if and only if the value of a quantity

[12] A profound impossibility theorem of non-contextual hv models is due to Gleason in 1957 ([57]) for Hilbert space of dimensionality greater than 2. Bell has a simplified version in his [13]. The conclusions are similar Kochen and Specker's in their [75].

[13] A simplified demonstration of Gleason's theorem can be found in [71, appendix A]. An enlightening presentation of the Kochen-Specker paradox can be found in [93, chapter 5].

measured depends either on which other quantities are measured, or on relevant features of the measuring apparatus. Note that there are two kinds of contextuality to distinguish in the definition above. The distinction was made, albeit not explicitly, by Bell.[14] Shimony has made the distinction explicit and dubbed them respectively *algebraic contextuality* and *environmental contextuality*. Here is how Shimony defines algebraic and environmental contextualities:

> [...] an "algebraic" context is one which specifies the quantities (or the operators representing them) which are measured jointly with the quantity (or operator) of primary interest, whereas an "environmental context" is one a specification of the physical characteristics of the measuring apparatus whereby it simultaneously measures several distinct co-measurable quantities.[15]

Other distinctions exist in the literature. Redhead distinguishes between *ontological contextuality*, *measurement contextuality* and *environmental contextuality*.[16] As far as we can tell, Redhead's ontological contextuality and measurement contextuality are both of Shimony's algebraic type: they both express the dependence of the result of the measurement of a given quantity on the other quantities measured jointly or considered alongside with this quantity. However, unlike Shimony's algebraic contextuality, they come with an interpretation of the observables and the measurement process. While both are of the algebraic type, Redhead's ontological and measurement contextualities differ from each other in that they come with two different interpretations. Ontological contextuality is algebraic contextuality equipped with an interpretational framework including possessed definite values before measurement and faithfulness of measurement. In such a framework, a measurement faithfully reveals the possessed value of the system. Outcomes depend on the algebraic context, because possessed values do. By contrast, measurement contextuality is also a kind of algebraic

[14] [13, p. 9].
[15] [104, p. 2].
[16] [94].

contextuality, but which comes with an interpretational framework in which the values of measurement correspond to relational properties. That is to say, a measurement does not reveal the possessed value but rather “a holistic relational attribute”, depending on both the system and the measurement quantum mechanical context.

Clifton and Pagonis distinguishes between two concepts of contextuality which they label $contextualism_1$ and $contextualism_2$.[17] $Contextualism_1$ is Shimony's environmental contextuality equipped with the interpretation that the values measured correspond to dispositions, that is, not categorical properties of the quantum system. They arise in an “unremarkable” way from both the initial state of the system and the experimental context. Thus, $Contextualism_1$ has no ontological significance. $Contextualism_2$, by contrast, corresponds to the causal dependence of possessed values on the context of measurement. Thus, $contextualism_2$ has strong ontological significance.

Shimony's and Redhead's environmental contextualities, despite sharing the same name, do not strictly overlap. Redhead makes it explicit that environmental contextuality includes only non-quantum mechanical aspects of the apparatuses. By contrast, Shimony's definition above seem to be completely general. If true, then Shimony's environmental contextuality is more general than Redhead's. In the following, we shall in general consider only Shimony's notions of contextuality.

With these clarifications above, we can now present the types of hv models we shall consider in this dissertation, i.e. stochastic, contextual, determinate-outcome hv models.

4.1.4 Stochastic models

Three different groups of hv models and corresponding derivations of the BI can be distinguished:[18]

---

[17] See [30].

[18] According to an important theorem by Arthur Fine, for each given experiment the three conditions above are equivalent (See [48], [49].) As said before, we shall analyze Fine's results in Chapter 5.

1. the outcomes can be accounted for in terms of a local deterministic hidden variable model,

2. the outcomes can be accounted for in terms of model with factorizable probability distributions,

3. the outcomes can be accounted for by a probabilistic model in which there are well defined joint-probabilities for incompatible observables.

The experimental results refute every set of assumptions that allows for the derivation of a Bell-type theorem.

In the rest of the dissertation, we shall mostly be concerned with sets of assumptions characterized by 2. above, that is with Bell-type theorems in which the BI are derived from a condition of factorizability on the probability distributions of the outcome-events. Such theorems constitute a generalization of the original version of Bell's theorem to the stochastic case. Such theorems apply to models in which the outcomes are determined only stochastically by the λ together with the experimental context. Let us say a little more about what it means for a probabilistic model in Bell-type situations to be stochastic.

First, not all probabilistic models are stochastic. Any determinate hv model of the Bell-type situations is a model where the complete state determines, deterministically or stochastically, and possibly together with the experimental context, the outcome of a given experiment. That said, while the system is taken to be really in one of the states $\lambda \in \Lambda$, we are in a state of ignorance of which one. Our ignorance about the specific state the system under consideration is in introduces a first layer of probabilities: any determinate model includes a probability measure on the possible λ. Any such model is thus a probabilistic model in the sense that it assigns probabilities to outcomes of Bell-type measurements. The probabilities for outcomes are obtained by averaging over the possible λ. The model will be satisfactory if the probabilities assigned are in agreement with the statistical distributions observed in experiment. This first layer of probabilities, the one representing our ignorance about the specific state the system under consideration is in, is independent of the issue of whether the model is deterministic or stochastic.

Contextual models entail a second layer of probabilities in the following way. In such models, the system under consideration is really in one of the complete states λ, but the λ do not alone determine the outcomes of a given experiment. Rather, they determine the outcomes together with the experimental context. Now, we may be ignorant of the exact state of the experimental context, or of how exactly it influences the outcomes. Typically, the model may take into account the possible influence of unknown and possibly uncontrollable features of the measurement apparatuses. In such a case, the λ do not determine an outcome bound to be obtained in a given experiment, but only prescribe probabilities of outcomes. The second layer of probabilities in some contextual models thus results from averaging over the unknown context. Here again, the fact that the model gives us only probabilities of outcomes does not imply that the model is not deterministic. Whether the λ and the experimental context determine the outcomes deterministically or stochastically still remains open. So, contextual hv models are perfectly compatible with determinism.

A hv model is not stochastic because it assigns probabilities to outcomes because of our ignorance of the underlying situation. Rather, it is stochastic if and only if the complete state of both system and experimental context still does not determine completely the result of a measurement. This typically occurs because a stochastic hv model includes a stochastic dynamics at the "hidden" level. A third layer of probabilities is thus introduced in such models. Stochastic models, even in the ideal case of a perfect knowledge of the underlying physical situation, would only assign probabilities to outcomes. Arguably, such models involve a form of irreducible indeterminism (which may concern the determination of the outcomes by the λ, by the experimental context, or both).

To sum up then, we have the following layers of probabilities:

1. Probability measure over the complete state – probabilities mirroring our ignorance of the underlying physical situation for the system – compatible with determinism;

2. Probability measure over the experimental context – probabilities mirroring our ignorance of the underlying physical situation for the apparatus(es) – compatible with determinism;

3. Probabilities of outcomes, given the λ and the complete state of the experimental context – incompatible with determinism.

Now that we have a clear idea of what it means for a model to be stochastic, we can turn to the issue of how the BI are derived from such stochastic models. As said above, stochastic hv models satisfy some BI if and only if they satisfy the so-called factorization condition, or Factorizability. In accordance with the above, the upshot of the experimental violation of the BI in these cases is that no factorizable hv model can give an account of all quantum phenomena. Let us say more about Factorizability.

4.1.5 Locality as Factorizability

Factorizability, or the factorization condition, is a condition on hv probabilistic models of the outcomes of the experiment. It is supposed to capture the locality part of Local Realism[19] within the stochastic context. The factorization condition is thus supposed to be a locality condition on determinate-outcome hv probabilistic models. Locality is here understood as the independence of a nearby outcome-event from remote events. The link between Factorizability and Locality is supposed to go both ways: on the one hand, violations of Factorizability amount to violations of Locality, and, on the other hand, with Factorizability satisfied, Locality is secured.[20]

[19]Definition 1, Subsection 4.1.1.

[20]Note that, in order to complete a Bell-type modus tollens argument, there is no need that satisfaction of Factorizability be a sufficient condition for Locality to be satisfied. What we only need is that a violation

Taking for example the case of a spin measurement, the rigorous formulation of factorizability is the following: Let λ stand for the "hidden state"[21], i for the setting of the apparatus **A** measuring the spin of one of the systems, j for the setting of the apparatus **B** measuring the spin of the other system[22], and a (respectively b) for the possible outcomes of measurements by apparatus **A** (respectively **B**) of the spin component in the direction defined by i (respectively j).[23] The probability for the outcomes a and b to jointly occur, conditional on i and j, and given the complete state λ, is simply the product of the probability of a, conditional on i by the probability of b, conditional on j:

Definition 5 – ***Factorizability***

A probabilistic model for Bell-type situations is factorizable if and only if:

$$p(a, b|i, j, \lambda) = p_(a|i, \lambda)p(b|j, \lambda).$$

The factorization condition[24] thus requires that the probabilities of outcomes at each end of the experiment be statistically independent, each conditional on the complete specification of the states and measurement protocol at its respective end of the experiment. To sum up then, stochastic Bell-type theorems express that any stochastic hv model that satisfies the factorization condition also satisfies some BI.

of Factorizability implies a violation of Locality. That is to say, it would suffice that Factorizability be a necessary condition for Locality. That said, the mainstream understanding seems to be that Factorizability is not only a necessary but also a sufficient condition of Locality.

[21] Whether λ is a single variable or a set of variables does not matter in the context of the following.

[22] The possibility that the apparatuses might also have hidden states will be discussed later in Section 4.3.

[23] In the following, we will denote the variables as capitals and the values of the variables as small letters. Thus, a and b are values of the outcome variables A and B; λ is a value of the complete state variable Λ.

[24] In our formulation of Factorizability, we take apparatus settings i and j as well as the hidden state λ as a variable on which we conditionalize, and not as a parameter which appears as an index. In the deterministic case, the settings of the apparatuses have to be taken as parameters on pain of contradiction (see [139, p.61-72] for more details). Since we confine ourselves with the stochastic case, we are not constrained to define a new probability measure where i and j are fixed. We shall indeed consider in general that the settings of the apparatuses are the result of a stochastic physical process, and not the result of the free choice of the experimenters. Now, whether the complete state λ should be taken as a parameter or a variable depends on whether or not we assume that we can assign a probability to it. We shall assume throughout that probabilities can be assigned both to λ and to i, j; therefore we shall consistently use the framework of conditional probabilities.

The interpretation of the experimental violation of the BI in this case amounts to interpreting the fact that no probabilistic model in which the probabilities for outcomes are factorizable in the above manner can give an account of Bell-type phenomena. This in turn has proved very difficult to analyze and has been the focus of foundational studies concerning Bell-type results for the past twenty five years.

4.2 The mainstream interpretation and its critics

4.2.1 The mainstream interpretation

Factorizability was introduced by Bell as a formal expression of a locality condition for probabilistic hv models. If that is the correct way to formalize the requirement of locality, then the conclusion of the Bell-type modus tollens is simply that no local realist model can give an account of all quantum phenomena. This is troubling because it seems to be in conflict with Relativity theory.[25] Most of the development in the analysis of the experimental violation of the BI by quantum mechanics and quantum phenomena have focused on the issue of whether such violation stands in conflict with the principles of Relativity theory.

An influential result in this regard was shown by Jarrett and Shimony (independently) in the mid-eighties. They showed that Factorizability divides up into two distinct conditions on the probability distributions of the outcome-events: Outcome and Parameter Independence.[26]

[25]This might hinge on a certain interpretation of Relativity theory. We shall not enter the debate over Relativity theory and its interpretations. Our aim here is to give an overview of the mainstream discussion in the Bell-literature. For a serious discussion concerning Quantum Mechanics, Bell-type theorems and the theory of Relativity, see [83] and references therein.

[26]The distinction between the two conditions was actually made for the first time by Suppes and Zanotti in [117]. Van Fraasen in [129] (reprinted as [130]) has used the same distinction for his own analysis. That said, Jarrett and Shimony were the first to claim that the distinction is relevant to the issue whether the experimental violation of the BI stands in conflict with Relativity theory. For the details, see [72] and [102], and also [73] and [103].

Note also that the phrases 'parameter independence' and 'outcome independence' are Shimony's. Jarrett calls similar conditions respectively 'locality' and 'completeness'. As Jones and Clifton notice in [74], Outcome Independence (henceforth OI) expresses the requirement that the outcome of the measurement on one subsystem be independent of the outcome of the measurement on the other one:

Definition 6 – ***Outcome Independence***

A probabilistic model for Bell-type situations satisfies Outcome Independence if and only if for all a, b, i, j, and λ:

$$p(a, b|i, j, \lambda) = p(a|i, j, \lambda)p(b|i, j, \lambda)$$

OI is a condition of statistical independence of the outcome-events. OI is not enough to obtain Factorizability because OI does not require the "separation" of the two factors i and j in the product of the conditional probabilities. In addition to mere statistical independence, what is needed to obtain the factorization condition is Parameter Independence (henceforth PI):

Definition 7 – ***Parameter Independence***

A probabilistic model for Bell-type situations satisfies Parameter independence if and only if for all a, b, i, i', j, j' and λ :

$$p(a|i, j, \lambda) = p(a|i, j', \lambda) \quad \textit{and} \quad p(b|i, j, \lambda) = p(b|i', j, \lambda),$$

that is, the requirement that the outcome of the measurement on one subsystem not depend upon the setting of the apparatus measuring the other one.

Shimony's and Jarrett's conditions are not identical, since the latter includes the hidden variables of the measurement apparatuses' states as factors, whereas the former does not. Jones and Clifton show that whether or not one considers these variables makes a difference to the issue of whether the violation of Bell-type inequalities by quantum theory and quantum phenomena stand in conflict with STR. We shall present their argument in Section 4.3.4 and deal more precisely with this issue in Chapter 7. For now, let us confine ourselves to Shimony's conditions.

That Factorizability is equivalent to the two above conditions taken together is *prima facie* an interesting result as far as the interpretation of the experimental violation of the BI is concerned. Granted, without any control of the hidden variables λ, we cannot test any of these two conditions independently of one another, so that we cannot determine empirically whether Bell-type systems violate one and not the other. That said, the distinction between these two conditions is interesting at the level of the interpretation. The distinction allows one to make the interpretation of the violation of BI by experiment more precise in narrowing down what one has to give up in any hv model for quantum probabilities and quantum phenomena. Indeed, the conclusion that any hv model must violate Factorizability entails that any hv model must fail to satisfy either PI or OI (or both). In other words, we may have to give up only one of the two.

Jarrett and Shimony have made the further claim that the distinction between PI and OI is relevant to whether the experimental violation of Bell-type inequalities by quantum theory and quantum phenomena stand in conflict with Relativity theory. They argued that, while any theory violating PI would stand in conflict with Relativity theory, a theory violating only OI could stand in "peaceful coexistence" with it.[27] The core of their argument for this is the claim that while a failure of PI could in principle lead to superluminal signaling, it is not the case for a failure of OI.

Here is how Jarrett and Shimony's argument goes.[28] Were we to have control over the hidden variables (that is, were we able to prepare and maintain a collection of quantum systems in the same complete state), then we would have the means to signal through failure of PI. Concretely, consider that the experimenter at one end, say the left, could prepare a collection of systems in a specific state, and maintain these in such a fixed state as well as

[27]The phrase is Shimony's in [100].

[28]See [72, p. 573-576] for the first formulation of the argument. The argument is completed in [101] and [73]. Our formulation of the argument is a synthesis of the series of argument formulated by Jarrett and Shimony.

control the settings of the apparatuses. If this were the case, the experimenter at the other (right) end could change the probabilities of outcomes on the left side by changing the setting on her (right) side, which would change the observable statistics on the other (left) side and thus constitute a case of signaling. By performing measurements on a large enough ensemble of Bell-type pairs of subsystems each spread over a large enough distance, one could thus efficiently send a signal at arbitrary speed. If the probability of the outcomes is independent of any parameters that could be controlled by the experimenters, no superluminal signaling can be achieved, even if the outcomes are correlated, i.e. even if OI fails. Finally, taking a ban on superluminal signaling as one of the principles of Relativity theory, Jarrett and Shimony conclude that our two best theories do not necessarily stand in conflict provided that we consider OI only, and not PI, as being violated by quantum theory and quantum phenomena.

Foundational studies on the matter have since mainly focused on the interpretation of PI, of OI, and of their respective potential failure. Most of the philosophers have followed Jarrett and Shimony on the argument above and have developed what we shall call a mainstream interpretation of Bell-type results. Let us start with violations of PI. The possibility of signaling is taken to be indicative of the existence of an underlying causal process. Because signaling can take place superluminally, the underlying causal process would be non-local. So, a violation of PI is taken to be indicative of the existence of a non-local causal process.

Still according to the mainstream interpretation, the argument above shows that a violation of OI is *not* indicative of a non-local causal process. Violations of OI are interpreted to be indicative of the fact that subsystems of the kind we are considering are not separate systems at all, despite their spatial separation. A violation of OI is thus interpreted as indicative of a form of holism.

We wish to emphasize that such an interpretation contains *three claims*. First, while PI is a locality condition in a sense connected with the theory of relativity, OI is not; second, that, while a violation of PI is indicative of a underlying causal influence, a violation of OI is not; and third, granted the first two claims, a violation of OI is interpreted as being indicative of a form of (not necessarily nonlocal and causal) holism.[29]

4.2.2 The mainstream interpretation challenged

So much for the mainstream interpretation. Let us now mention that, since Jarrett's and Shimony's seminal papers in the mid-eighties, there has been an ongoing, albeit rather underground, controversy over the interpretation of Jarrett and Shimony's analysis of Factorizability. Let us lay down some of the contributions in the literature that have denied one or several of the three claims above. Earman has denied that factorizability appropriately captures notions of locality and/or action at a distance.[30] Butterfield has argued that a ban on superluminal causation does not favor PI over OI.[31] Jones and Clifton have argued that, under some specific circumstances, a violation of Jarrett's version of OI can lead to superluminal signaling.[32] Maudlin argues that Jarrett's and Shimony's distinction is not relevant as an analysis of the underlying causal picture.[33] Finally, Berkovitz argues that, while the distinction is relevant as far as superluminal signalling and the conflict with Relativity theory are concerned, it is not concerning the issue of action at a distance, superluminal causation and holism.[34]

[29] For an overview of the interpretation of PI, OI and of their potential failure, see [31], or, more recently, [23]. The literature on the interpretation of the experimental violation of the BI in terms of holism is plentiful. Howard's seminal paper ([69]) was indeed followed by, among others, Teller ([123], [124]), Howard himself ([70]), Healey ([60], [61], [62]), and Dickson ([35]). The debate is still vivid, as the paper by Winsberg and Fine ([135]) and the answer Fogel gives to it ([54]) prove. A fairly recent synthesis on these debates can be found in [63]. Another influential (and not unrelated) trend for the interpretation of OI and its failure, is Shimony's and Redhead's interpretations in terms of non-causal relation or 'passion-at-a-distance'. See for example [94] for an argument to the effect that passion-at-a-distance differs from action-at-a-distance in terms of the 'robustness' of the influence.

[30] [41].

[31] [22].

[32] [74].

[33] [83, p.94-95].

[34] [17, 18].

We shall undertake a systematic examination of the mainstream interpretation. We rigorously assess whether or not there are precise frameworks for discussing locality and causality in which the three claims which we said constitute the mainstream interpretation hold. Let us recall these three main claims:

Definition 8 – ***(PI-LOC)***:

Parameter Independence is a locality condition while Outcome Independence is not – locality being understood in a sense linked with Relativity Theory.

Definition 9 – ***(¬PI-CAUS)***:

A violation of Parameter Independence is indicative of an underlying causal relationship, while this is not the case for a violation of Outcome Independence.

Definition 10 – ***(¬OI-HOL)***:

Given that (PI-LOC) and (¬PI-CAUS) hold, a violation of Outcome Independence is indicative of a form of non-separability or holism.

In the following section, we shall explain how Jarrett and Shimony's argument in terms of in principle superluminal signaling fails to support (PI-LOC), (¬PI-CAUS), and hence (¬OI-HOL). In doing so, we shall dwell on arguments formulated by some of the critics of the mainstream interpretation. In particular, we shall rely on Jones and Clifton's (1993).[35] The upshot will thus be that the mainstream interpretation lacks argumentative support.

Before turning to this, we need to make one note about our notation. As briefly mentioned before, Jarrett's and Shimony's conditions differ from each other. More precisely, they differ from each other exactly in that Jarrett includes the hidden variables of the apparatuses as factors, while Shimony does not. We shall soon argue (Section 4.3) that a complete picture of the Bell-situation should include the hidden variables of the apparatuses. So, our analysis

[35] [74].

utilizes Jarrett's conditions, which he called Locality and Completeness, and not PI and OI, which are Shimony's corresponding conditions without the hidden parameters of the apparatuses. While we favor Jarrett's conditions, we much prefer Shimony's terminology over Jarrett's, because it is more neutral. Thus, from now on, we shall stick to Shimony's terminology, but in order to denote Jarrett's conditions. We shall add stars to the acronyms every time we do so. So, OI* and PI* are Outcome Independence and Parameter Independence in which the hidden variables of the apparatuses are taken into account. Similarly, (PI-LOC)*, (¬PI-CAUS)* and (¬OI-HOL)* are the theses corresponding to (PI-LOC), (¬PI-CAUS) and (¬OI-HOL), when the hidden variables of the apparatuses are taken into account. With this in mind, let us now turn to show that Jarrett and Shimony's argument fails to satisfactorily support the mainstream interpretation.

4.3 How Jarrett and Shimony's argument fails

In this section, we shall recall Jones and Clifton's arguments to the effect that Jarrett's and Shimony's argument in terms of superluminal signaling does not satisfactorily support either (PI-LOC)* or (¬PI-CAUS)*[36]. We shall then notice that it does not follow from Jones' and Clifton's argument[37] that either (PI-LOC)* or (¬PI-CAUS)* is false. We shall conclude that the question of whether (PI-LOC)* or (¬PI-CAUS)* hold can be settled only if a precise definitions of locality and causality (respectively) are adopted. A clear definition of a specific spacetime framework and of locality in that framework are necessary to discuss (PI-LOC)*. A specific theory of causation is necessary to discuss (¬PI-CAUS)*. The remaining chapters show that providing rigorous frameworks indeed proves to be useful.

Jones and Clifton develop their analysis on the basis of Jarrett's conditions, conditions

[36]Note that Kronz, in [77], had leveled a similar objection against Jarrett and Shimony. Jones and Clifton improve on Kronz in so far as they show that the objection holds only in the case where one takes into account the hidden variables of the apparatuses' states.

[37]Note that they do not pretend it does.

which include the so-called hidden variables of the apparatuses' states as factors. We shall now discuss the distinction between Shimony's and Jarrett's conditions, and explain why the latter are more interesting.[38]

4.3.1 Surface correlations and the underlying causal picture

Bell-type experiments typically amount to the observation of correlations between observables outcomes. How do we reason when facing correlated events? Generally, we tend to think of correlations as indicative of an underlying causal relationships. This in turn hinges on the idea that probability distributions over events and causation go hand in hand. Where does such an idea come from? It comes as a development of regularity theories of causation. Regularity theories are a very natural type of theory of causation, in the sense that they give a simple empirical criterion to recognize causal relationships. According to such theories, a cause is consistently followed by its effect. Hence, whenever one observes that an event e (e.g. my hand in fire) is consistently followed by an event f (e.g. my hand burns), one has a good indication that e is a cause of f. This, of course, is far too rough, and demands elaboration to qualify as a satisfactory criterion for causal processes. Without entering in any details for now, let us only sketch how the theories of probabilistic causation come as a development of the intuition behind regularity theories of causation.

One starts by noting that the criterion for qualifying as a cause given above fails in the case of imperfect regularities between a cause and its effect. Imperfect regularities may appear in the two following cases: 1. e is not always followed by f because the occurrence of the effect depends on factors other than e – e is a deterministic but partial (maybe necessary but not sufficient) cause of f; 2. e is not always followed by f because the world is fundamentally and irreducibly indeterministic. Theories of probabilistic causation mainly

[38] Jones and Clifton confine themselves to notice the distinction between Shimony's and Jarrett's conditions and draw the consequences in case one takes Jarrett's view. They do not take a stance over which view is preferable.

aim at accommodating these two cases.

A central idea of theories of probabilistic causation is that the occurrence of a cause increases the probability that the effect occur.[39] Such an idea can be construed either within the framework of counterfactual conditionals, in which the consequent is a proposition assigned a probability, or in using the framework of conditional probabilities. Construed in terms of counterfactuals, the idea amounts to say that e is a cause of f just in case the probability that f occurs is higher if e occurs than if e did not occur. Formally:

Definition 11 – ***Causation: counterfactual conditional****:*

In a counterfactual conditional account, e is a cause of f if and only if, for $x \in [0,1]$,

$$(e \quad occurs) \rightarrow p(f) = x \quad and \quad \neg\,(e \quad occurs) \rightarrow (p(f) \leq x)$$

Construed in terms of conditional probabilities, the same idea amounts to say that the probability of f, conditional on the occurrence of e, is higher than the probability of f, conditional on the non-occurrence of f.

Definition 12 – ***Causation: conditional probabilities****:*

In a conditional probabilities account, e is a cause of f if and only if $p(f|e) \geq p(f|\neg e)$

In this dissertation, we shall not deal specifically with counterfactual theories of causation. We shall instead focus on the theory of probabilistic causation which uses the formalism of conditional probabilities. This type of framework is the most commonly used in the Bell-literature. Accordingly, we shall consider the completed state λ and the settings i and j as potential causal factors on which to conditionalize and not as fixed parameters. This entails that λ, the settings i and j, as well as all other factors, be construed as events which can be ascribed some probabilities of occurrence.

[39]Probabilities have to be interpreted as objective in such theories. That said, this does not involve that such objective probabilities be interpreted as irreducible chances.

There are convincing arguments that Definition 12 fails to characterize causes. Indeed, there are well known counterexamples to the claim that every increase in probability, as in Definition 12, is indicative of a causal relation[40] and also to the claim that every causal relation results in an observable increase in probability.[41] Probabilistic conditions must be supplemented by other conditions such as manipulability or continuity in spacetime to characterize causal processes. We shall not dwell much on this for now. We provide a more complete presentation of theories of probabilistic causation in Chapter 7.

For the moment, let us illustrate how the central idea of theories of probabilistic causation seem to give a natural account of imperfect regularities by way of a few examples. The main idea is the following: imagine that all we have is some probability distributions over events; one then hopes that the theory allows one to infer the *underlying causal picture*, which includes the *potential causal factors* as well as the *various causal relationships* between these factors.

An example is the following. Imagine that you notice that you have a higher chance of suffering from stomach pain after eating fast food than from eating a meal in accordance with the USDA nutrition recommendations: you will probably find yourself inferring that high fat / high sugar food causes indigestion. This would be a case of direct, albeit either partial or indeterministic, causation. Now, let us turn to two more elaborate examples.

First, consider that your two best friends, Jules and Jim, are both somewhat cyclothymic, that is, they are, from days to days, either in an excellent or in a execrable mood. Now imagine that you note that, while their respective "mood-states" are randomly distributed over the days, one is always in the opposite state of the other. That is, whenever Jules is in a good mood, Jim is in a terrible one, and vice versa. Now if you further know for sure that your two best friend do not know each other, you will probably find yourself quite

[40] A typical counterexample of such spurious correlations are common cause patterns.

[41] Typical counterexamples of cancellation of correlations are related to reversal of probabilities as in Simpson's paradox.

puzzled by such correlations. You will probably feel satisfied once you realize that 1. both Jules and Jim go every morning to the same bakery to buy some croissants for breakfast; 2. while Jules love the crispiness of a thoroughly cooked croissant, Jim appreciates much more the sensation of an almost uncooked croissant softly melting in his mouth; 3. the baker at their common bakery randomly under- or overcooks his croissants. Based on a pretty crude theory of psychology (reducing the mood-state of Jules and Jim to an effect of how much they enjoy their breakfast), you have a satisfactory scheme of explanation, or underlying causal picture, for the correlations observed. The perfect correlations between two events, none of which causes the other, are thus explained in terms of the random duration of baking of the croissants in the common bakery.[42]

Second, consider that another friend of yours, June, suffers from incurable anxiety. To cope with such anxiety, she took up smoking very early in her life. This, according to all your knowledge, increases her chances to die from heart attack. Imagine now that, smoking being not enough, June starts each of her days with a 6 mile run. And imagine further that, feeling exhausted after a day of work, she takes a glass of red wine in order to relax every night. It might well be the case that June does not have higher chances of having a heart attack than the average non-smoker. Even if a very few people would deny that smoking is a cause of heart attack, the corresponding rise in probability is canceled in the case of June by other factors. Indeed, regular exercise strengthens the heart while a regular consumption of red wine lowers cholesterol. Again, were you to know only about June's heavy smoking, you would be surprised that her chances of having a heart attack are not higher that an average non-smoking person. That said, you will probably feel satisfied when you discover about her other habitual means to fight anxiety. You have now a satisfactory scheme of explanation, or underlying causal picture for the correlations observed (or the

[42]Such treatment of correlations as arising from either a direct causal relationship or from two parallel causal path originating in a common cause was famously formalized for the first time by Reichenbach in [43]. We shall give a brief presentation of the Principle of Common Cause by Reichenbach in Chapter 6.

lack of them). Theories of probabilistic causation aim just at giving an empirical criterion to infer such causal picture, which includes the causal factors and the causal relationships between these factors, underlying some events, on the basis of the probability distributions over these events.

A final note about the theories of probabilistic causation. Causal relationships are not always positive. Here is an example of a non-positive causal relationship: regular exercise decreases the chances of obesity. Despite such decreasing effect, one might want to consider regular exercise as a causal factor in the causal picture. One would only have to make precise that the causal relationship between regular exercise and obesity is of a negative, or impairing, type. Arguably, our framework should accommodate this type of cause as well. So, it may be useful in some cases to modify the central idea that a cause increases the probability of its effect to: a cause changes the probabilities of occurrence of its effect, either positively or negatively:

Definition 13 – ***Causation: conditional probabilities* – *refined***:

In a refined conditional probability account, e is a causally relevant to f if and only if $p(f|e) \neq p(f|\neg e)$

In the examples above, all factors (the fat and sugar in the fast food, the amount of cooking the croissants received, and June's habits of smoking, running and drinking red wine) were observable. What the user of the a theory of probabilistic causation wanted to do is to figure out the relations between the causal factors and the events to be explained. What happens when no potential causal factor is observable? In such cases, one hypothesizes an underlying, even if not directly observable, causal picture behind the phenomena.[43] Thus, to

[43]Whether one wants to take a realistic stance on such a causal picture, or prefers to remain agnostic as to the ontological status of the causal factors and causal relationships postulated remains open here. In the latter case, that is, if one wants to remain agnostic as to what the underlying causal picture *is*, it is still interesting to try to figure out what the underlying causal picture *could be*. In particular, it is interesting to figure out what causal pictures are compatible with the phenomena and/or with our best theories.

the “surface description” of the phenomena – the observed correlations and the controllable factors, one adds the description of the *hidden, i.e. unobservable and possibly uncontrollable, causal picture* underlying the surface phenomena[44]. Arguably, it is more difficult to decide what should count as a causal factor and what should not in this case.

The type of philosophical analysis that has been applied to Bell-type situations fits into such a heuristic schema: one aims at giving the description of the hidden causal picture, which includes a set of hidden causal factors and causal relations between such hidden causal factors and outcome events. In introducing the λ in the factorization condition, Bell introduces a causal factor which could cause correlation of events as a result of two parallel causal chains originating in a common cause (like the baking time in the example of Jules and Jim). Jarrett and Shimony’s analysis of the factorization condition can be seen as of the same kind. It consists in distinguishing between the potential relevant paths underlying the correlations. One such path goes from the parameter on one side to the outcome on the other side. The other path goes between the outcomes. We thus end up with three possible causal paths underlying the observed correlations, as shown in Figure 4.1.

In Shimony’s analysis, the hidden causal picture for Bell-type situations includes: the settings of the apparatuses i and j, the outcome-events a and b, and the state λ. However, in such a picture, one element of the experiment is indeed generally ignored, namely the hidden variables of the states of the apparatuses. These elements should enter into the picture and be distinguished from the surface states of the apparatuses as well as from the hidden variables of the system.

[44]See van Fraassen in [130], 103 sq., for a more systematic distinction between the surface and the hidden description of a phenomenon.

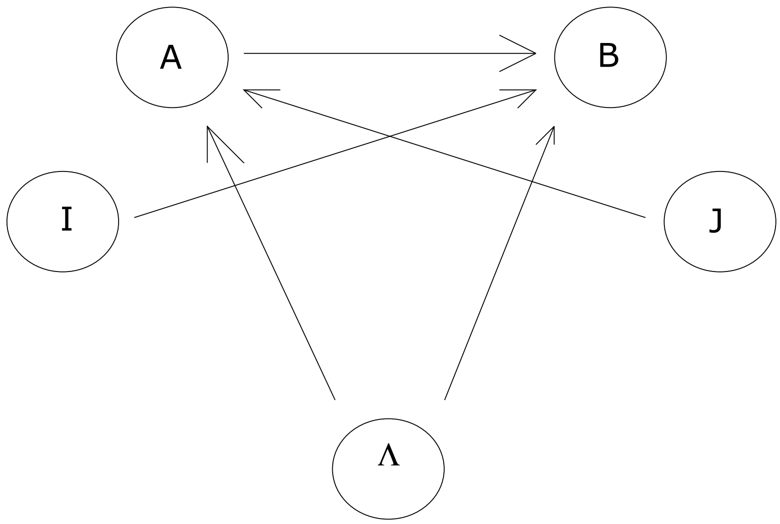

Figure 4.1: A traditional picture of the putative causal structure underlying Bell-type phenomena

4.3.2 From Shimony to Jarrett: A completed causal picture

Bell was one of the first to express his interest in taking into account the hidden variables of the apparatuses' states in his [11]. In his paper, Bell writes:

> Actually we can, and we should, be somewhat more general. The instruments themselves could contain hidden variables which could influence the results. ([11], 37)

In the literature, not much attention has been paid to such a generalization of Bell's result to the case where hidden variables for the measurement apparatuses were taken into account. In general, the possible relevance of the hidden variables of the apparatuses' states is ignored.

A rather simple argument in favor of taking the hidden variables of the apparatuses into account is the following. In constructing a causal picture underlying some phenomena, we should aim at completeness. In the case in which we try to design the *hidden* causal picture underlying some phenomena, as in the case of Bell-type situations, then no possibility should be left aside, since no empirical access is allowed. Therefore, there is no a priori reason to leave out the possibility that the hidden variables of the apparatuses may be relevant in Bell-type situations. They could enter into the hidden causal picture underlying the experiment.

We have already mentioned that Jarrett's and Shimony's conditions differ exactly in that the latter does not include the hidden variables of the apparatuses' states, while the former does. Jarrett considers that the i and j are the "detector-states, including the switch setting and whatever else may be relevant.[45] By contrast, Shimony makes it explicit that the i and j can include various characteristics of the apparatuses, "but since the dependence of the result upon the microscopic features of the apparatus is not determinable experimentally, only the macroscopic features of the apparatus (such as the orientation of the polarization analyzers), in their incompletely controllable environment need be admitted in practice."[46]

Shimony's argument to *not* consider the hidden variables of the apparatuses' states as potential causal factors is that they are unobservable and uncontrollable in practice. By contrast, i and j are the parameters that are controllable in practice by the experimenters. All such controllable parameters are included in i and j, while everything else is left out. The reason for this is that Shimony is interested in the practical possibility of signaling. From the point of view of providing the hidden causal picture underlying the experiments, however, this is not enough. When introducing the hidden state λ in the first place, we introduce a non-observable and perhaps uncontrollable factor as well. If the possibility of empirically determining the dependence of the outcome on a variable X is necessary to consider X as qualifying as a causal factor in the causal picture underlying the phenomena, then λ does not qualify either. The point is that, as we explained above, the analysis of the Bell-type situations consists in supplementing the surface description of the phenomena,

[45][72], [73, p.65].
[46]Shimony, [102], [104, p.5].

which includes the outcomes as well as the macroscopic and/or controllable variables of the apparatuses' states, with a description of the hidden causal picture, which postulate potential causal factors and causal relationships. In such a hypothetical inquiry, none of the hidden causal factors are supposed to be either empirically observable or directly controllable. Because we do not have any direct access to the causal factors, we are confined to the level of speculation. That said, it seems reasonable to require that the causal picture be as complete as possible. So, in this perspective, and contra Shimony, we take that the hidden features of the apparatuses should be considered.

Moreover, contra Jarrett, we would maintain that the hidden variables of the apparatuses should be distinguished from the surface (observable and perhaps controllable) features of the apparatuses. Jarrett includes under the same factor the hidden features together with the settings and other surface features, in short: all that could be relevant to the result. We do not see any a priori reason to believe that there is a one-one correspondence between the hidden and the surface features of the apparatuses. In the case one considers that the surface features of the apparatuses are chosen by the experimenters, there might be various hidden physical situations underlying the same setting. In the case where the macroscopic features, in particular the settings of the instruments, are the result of (irreducibly indeterministic) stochastic processes, then there is every reason to believe that different settings can result from the same underlying physical situation. Second, the hidden features of the apparatuses may have an influence on the results through a causal path that does not go through the surface features. Finally, in keeping the hidden and surface features of the apparatuses separate, we keep separate the parameters that are controllable in practice from the ones that are controllable in principle only, a distinction which is certainly of interest. One could object that hidden features presumably influence surface features at the same end of the experiment. So, were we to have complete over hidden features, then surface features would become redundant. The interest of including the surface features as independent of the hidden features resides in that surface features could be changed in practice, but that change in the macrostate would only constrains the microstate to be one of the many compatible with the given macrostate. For all these reasons, we prefer keeping the hidden states of the apparatuses separate from their surface states, with the assumption that the hidden features of an apparatus do not influence the surface features of the same apparatus.

Thus, we shall consider in the following that a complete picture of the underlying physical situation includes the hidden variables of the apparatuses' states as potential causal factors, distinct from both the hidden variables of the system's state and the surface variables of the apparatuses' state. Let us then give a more complete specification of the states of the apparatuses by supplying i and j with the hidden variables γ and δ respectively. By including theses variables into the picture, we complement the causal picture of the experiment in the following way. We are now considering something like Figure 4.2:

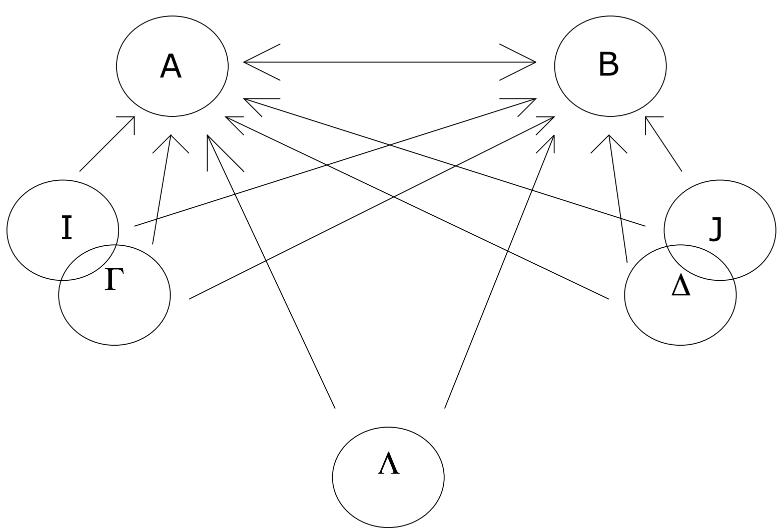

Figure 4.2: A completed picture of the putative causal structure underlying Bell-type phenomena

Further, the two conditions which together entail some BI now take he following form:

Definition 14 – ***Outcome Independence*:***

A probabilistic model for Bell-type situations satisfies Outcome Independence if and only if for all a, b, i, j, λ, γ, and δ:*

$$p(a, b|i, j, \lambda, \gamma, \delta) = p(a|i, j, \lambda, \gamma, \delta)p(b|i, j, \lambda, \gamma, \delta);$$

Definition 15 – ***Parameter Independence*:***

A probabilistic model for Bell-type situations satisfies Parameter Independence if and only if for all a, b, i, i′, j, j′ λ, γ, γ′, δ and δ′:*

$$p(a|i, j, \lambda, \gamma, \delta) = p(a|i, j', \lambda, \gamma, \delta'), \quad \textit{and} \quad p(b|i, j, \lambda, \gamma, \delta) = p(b|i', j, \lambda, \gamma', \delta).$$

Taking into account the hidden variables of the apparatuses as distinct potential factors allows us to analyze Factorizability as the combination of more than two conditions. Indeed, as new causal factors being considered, new causal paths also enter into the picture.

4.3.3 Hidden Parameter Independence: Local and Global

We now need to define the various probability conditions that correspond to the introduction of the hidden variables of the apparatuses' states as potential factors in our hidden causal picture. That is to say, we have to introduce what we shall call *Hidden Parameter Dependence*, for it corresponds to the statistical dependence of the outcomes over such hidden variables of the apparatuses' states.

Let us first note that, within the terminology of the Bell literature, Hidden Parameter Dependence is a form of contextuality. Given the various notions of contextuality in the literature [47], let us spend a little time explaining what kind of contextuality Hidden Parameter Dependence is. First, within Shimony's distinction between algebraic and environmental contextualities, Hidden Parameter Dependence is of the environmental type: it corresponds

[47] See Section 4.1.3.

to the specification of the physical characteristics of the measuring apparatuses, not to what other measurements are performed. Second, Hidden Parameter Dependence remains neutral as to what interpretation is made regarding the properties of quantum systems, whether they are categorical, dispositional or relational. Finally, Hidden Parameter Dependence does not include any constraint that the hidden variables of the apparatuses' states be strictly quantum mechanical. In short then, Hidden Parameter Dependence singles out the potential statistical relevance of the hidden variables, by contrast to the surface features, within the environmental context.

In the case of two-winged Bell-type situations, Hidden Parameter Dependence naturally divides into two different conditions, which we shall call Local Hidden Parameter Dependence and Global Hidden Parameter Dependence.[48] Consistent with what was said before, such a division corresponds to the distinction between the various causal paths through which, in a Bell-type situation, the hidden states of the apparatuses could influence the outcomes.

Local Hidden Parameter Independence (henceforth LHPD) expresses the idea that the outcome distribution at one end of the experiment is independent from changes in the hidden variables of the apparatus at this end. For example at the **A**-end:

Definition 16 – ***Local Hidden Parameter Independence****:*

A probabilistic model for Bell-type situations satisfies Local Hidden Parameter Independence at the **A***-end if and only if for all* a, b, i, j, λ, γ, γ', δ*:*

$$p(a|i, j, \lambda, \gamma, \delta) = p(a|i, j, \lambda, \gamma', \delta);$$

and similarly for the **B**-end.

Global Hidden Parameter Independence (henceforth GHPI) expresses the idea that the outcome distribution at one end of the experiment is independent from changes in the hidden

[48]What we call Local Hidden Parameter Independence corresponds to a violation of Jones and Clifton's Measurement contextuality (MC). What we call Global Hidden Parameter Independence corresponds to their Constrained Locality (CLOC).

variables of the apparatuses at the other end. For example, at the **A**-end:

Definition 17 – ***Global Hidden Parameter Independence****:*

A probabilistic model for Bell-type situations satisfies Global Hidden Parameter Independence at the **A***-end if and only if for all* a*,* b*,* i*,* j*,* λ*,* γ*,* δ *and* δ'*:*

$$p(a|b, i, j, \lambda, \gamma, \delta) = p(a|b, i, j, \lambda, \gamma, \delta');$$

and similarly for the **B**-end.

It is important to note that, in the expression for Global Hidden Parameter Independence for the outcome a, conditionalization is made on the outcome b. The point is, again, to try to differentiate between the various possible causal factors and causal paths underlying the correlations. GHPI is formulated so that its violation corresponds to the idea that a change in the hidden variables of the apparatus at one end is *directly relevant* to the outcome event at the other end. That is to say, it is supposed to correspond to a direct causal path between the hidden variables of the apparatus at one end and the outcome at the other end. Now, the probabilistic condition would not be appropriate if it did not exclude those cases where the relevance of the hidden variables of the apparatus at **B** on the outcome at **A** is obtained by Local Hidden Parameter Dependence at **B** together with Outcome Dependence between a and b. Conditionalizing on b is just what is needed to exclude these cases and separate the two causal paths, as shown in Figure 4.3.

With these condition in hand, we can lay out the relationships between Jarrett's and Shimony's conditions. With OI standing for Outcome Independence, OI* for Jarrett's "Completeness", PI for Parameter independence, PI* for Jarrett's "Locality" and, again, LHPI and GHPI respectively for Local and Global Hidden Parameter Independence:

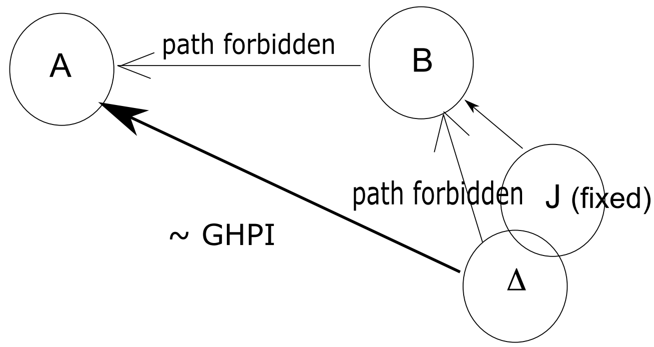

Figure 4.3: GHPI singles out the direct causal path from the hidden parameter at one end to the outcome at the other end, excluding an indirect causal influence the outcome variable B.

$$
\begin{aligned}
&(a) \quad PI^* \rightarrow PI + GHPI;\\
&(b) \quad PI + LHPI + GHPI \rightarrow PI^*;\\
&(c) \quad OI^* + LHPI + GHPI \rightarrow OI;\\
&(d) \quad OI + LHPI + GHPI \rightarrow OI^*.
\end{aligned}
$$
[49]

We introduced two new probability conditions for the analysis of the causal picture underlying Bell-type situations. Our justification for this has been a requirement of completeness regarding both the various potential causal factors and the various causal paths for the correlations. It remains to show how Jarrett and Shimony's argument fails to support either (PI-LOC)* or (¬PI-CAUS)*.

4.3.4 OI*, PI*, and Superluminal Signaling

The reason why Jarrett and Shimony's argument in terms of superluminal signalling fails to support either (PI-LOC)* or (¬PI-CAUS)* is simple. Jarrett's and Shimony's argument in favor of the claim that a violation of PI* is indicative of a non-local causal influence while a violation of OI* is not, is that the former can lead to in principle superluminal signaling while the latter alone cannot. Given an association between signaling and causation, they deduce that a violation of PI* is indicative of an underlying causal relationship while a violation of OI* is not. Given the experimental set up, they conclude that such a causal relationship is also non local.

Jones and Clifton have challenged the reasoning above. They show that certain cases of superluminal signaling can be reduced to a causal influence between outcomes alone. Hence, PI* and OI* do not single out specific causal paths as far as superluminal signaling is concerned. Hence, neither (¬PI-CAUS)* nor (PI-LOC)* are supported by the argument in terms of superluminal signaling.

Let us give a more precise account of Jones and Clifton's proof. They show that if one takes into account the hidden variables of the apparatuses' states, then a failure of OI*, together with a failure of Local Hidden Parameter Independence, imply a failure of PI* itself, which we know can lead to signal non-locality in principle. At first sight, such result could appear irrelevant to the question whether or not (¬PI-CAUS)* and (PI-LOC)* hold. Indeed, the mainstream interpretation is not endangered by cases in which failure of OI* along with failure of PI* can lead to superluminal signaling. Instead, the mainstream interpretation is that failure of OI* *alone* does not allow for superluminal signaling, and hence is not indicative of a non-local causal process. However, a closer look at Jones and Clifton's result allows us to see why it is relevant. In the cases of superluminal signaling they consider, they rule out the apparatus-outcome causal path by imposing GHPI.[50]

[50]See the proof in [74, p.303].

Theorem 1 – ***Jones and Clifton***:

$$\neg OI^{*} + \neg LHPI + GHPI \rightarrow \neg PI^{*}$$

Note that on the right hand side of the implication Local Hidden Parameter Independence fails, but Global Hidden Parameter Independence holds. It is interesting to understand why the failure of GHPD is required to entail ¬ PI*. In securing GHPI, Jones and Clifton aim at singling out the causal path $\neg LHPI + \neg OI*$ under the assumption that the possibility of signaling is characteristic of a causal path. We shall emphasize the importance of this assumption at the end of this section. So, the possibility of superluminal signaling obtained under the right hand side conditions is *not* obtained through the causal path that PI* aims at characterizing, i.e. the causal path between the state of the apparatus at one end and the outcomes at the other. Inclusion of GHPI serves to block the objection that, since the result shows only that a failure of OI* supplemented with failure Local Hidden Parameter Independence implies a failure of PI*, we are, in this case of signaling, using again the causal path corresponding to the failure of PI*, and not the one corresponding to the failure of OI*. Contra such an objection, the conditions above render the causal path between the apparatus state at the **B**-end and the outcome at the **B**-end unusable thanks to GHPI. Figure 4.4 should help to make this point clear.

Under the assumption of failure of Local Hidden Parameter Independence, an argument can be designed to the effect that a violation of OI* could be used in principle for superluminal signaling. Such argument is strictly similar to the one formulated by Jarrett and Shimony in the case of PI*. Indeed, imagine that experimenters had sufficient control on the hidden variables λ, γ and δ, such that 1. they prepare a set of systems all in the same state λ, 2. the experimenter ($E_{\mathbf{A}}$) at the **A**-end of the experiment sets all her measuring apparatuses to (i, γ), and 3. the experimenter ($E_{\mathbf{B}}$) at the **B**-end chooses to set her measuring devices sometimes as (j, δ), sometimes as (j, δ'). If this were the case, ($E_{\mathbf{B}}$) could change the proba-

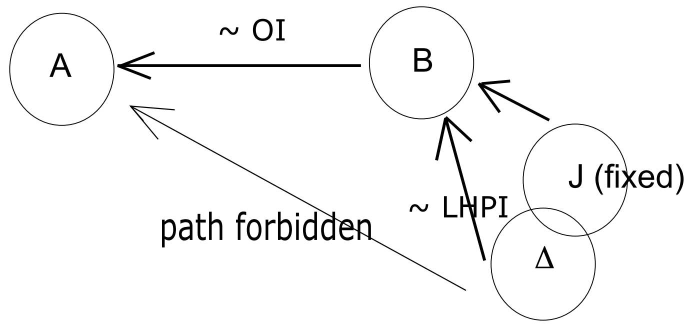

Figure 4.4: The causal structure underlying Jones and Clifton's theorem: the causal path corresponding to PI* is forbidden.

bilities of outcomes on the **A**-side by changing the setting on her side, which would change the observable statistics for ($E_{\mathbf{A}}$) and thus constitute a case of signaling. By performing measurements on a large enough ensemble of Bell-type pairs of subsystems each spread over a large enough distance, one could thus efficiently send a signal at arbitrary speed. Thus one can obtain in principle superluminal signaling using a violation of OI*.

4.4 Outline of part II

4.4.1 Strategy

Jones and Clifton have shown that, given that Local Hidden Parameter Independence fails, some cases of superluminal signaling are obtained in using a causal path between the outcomes. If true, then Jarrett and Shimony's argument in terms of superluminal signaling is not satisfactory as an argument for either (¬PI-CAUS)* or for (PI-LOC)*, so that (¬OI-HOL)* is in turn ill-motivated. If one takes Relativity theory to ban spacelike causal influences, then the traditional interpretation that it is preferable to give up OI* than PI* as far as the compatibility of the models with Relativity theory is concerned is ill founded.

Note that Jones and Clifton's result does not imply that either of the claims constituting the traditional interpretation, that is, (¬PI-CAUS)*, (PI-LOC)*, or (¬OI-HOL)*, is false. Concerning (¬PI-CAUS)*, that it is impossible to signal thanks to a violation of OI* alone is not sufficient to conclude that there is no causal relationship between the outcomes underlying the violation of OI*. Moreover, superluminal signaling thanks failure of PI* do not necessarily correspond to usage of a causal path between the parameter at one end and the outcome at the other end. The upshot is that the mere possibility of signaling using some correlations between some events, without any further refinement, is neither necessary nor sufficient a criterion for determining whether or not the correlations considered correspond to a causal path.

What is needed in order to assess whether or not (¬PI-CAUS)* holds is a more rigorous theory of probabilistic causation. We can see two options here. Either one attempts to refine the criterion in terms of signaling – as do the Manipulability Theories of Probability Causation (MTPC), or one rejects the criterion of manipulability and proposes another one. In the later case, there are two alternative theories of causation available on the market. One such theory is the counterfactual theory à la Lewis. Jeremy Butterfield has proved that (¬PI-CAUS)* does not hold within that theory.[51] We consider the last alternative available, namely the theories of causation as continuous processes in spacetime à la Salmon and Dowe. Chapter 7 is devoted to assessing whether or not (¬PI-CAUS)* holds in MTPC and/or STPC.

Concerning (PI-LOC)* now, we find ourselves in an embarrassing situation. On the one

[51] [22].

hand, if one supposes that locality reduces to a ban superluminal signaling, then PI* is a condition of locality while OI* is not and (PI-LOC)* holds. On the other hand, if one takes that locality reduces to a ban on spacelike causal influences, then OI* and PI* in general both qualify as locality conditions and (PI-LOC)* does not hold. What is needed in order to settle the situation is a rigorous definition of locality. Arguably a definition of locality, hinges on the space-time structure in which it is considered. It seems then reasonable to require that discussions about locality take place within a clear space-time framework. We provide one such rigorous spacetime framework in the next chapter (Chapter 6) in which it is possible to show that (PI-LOC)* holds.

4.4.2 Outline

In Chapter 6, we shall turn to the issue of locality. In Jarrett's and Shimony's argument, as in most of the Bell literature, no space-time structure is specified in any rigorous way.[52] We shall begin by correcting this problem. We define a precise spacetime framework for the discussion of locality, in which we provide a rigorous argument in favor of (PI-LOC)*. To do so, we shall extend Earman's argumentation from his 1987 paper.[53] In this paper, Earman does not deal with PI and OI, but only with Factorizability. Further, Earman confines his discussion to the deterministic case. We extend Earman's discussion in two ways: first, we discuss Outcome Independence* and Parameter Independence* and not only Factorizability. Second, we generalize the discussion to the stochastic case. We shall show that a rigorous argument in favor of (¬PI-LOC)* can be made within our framework. We shall lead a discussion concerning how to generalize Einstein Locality to the stochastic case. With this notion of Stochastic Einstein Locality (SEL) in hand, we shall argue that SEL entails the space-time version of PI* but not the spacetime version of OI*. Instead, the requirement that

[52]Exceptions are [64], [41], [21], [83] and [17, 18].
[53][41].

spacetime version of OI* holds is shown to be equivalent to the demands of the spacetime version of the Principle of Common Cause when applied to Bell-type situations.

In chapter 7, we finally turn to examining the claims of the mainstream interpretation concerning causal links: (¬PI-CAUS)* and (¬OI-HOL)*. We consider two theories of probabilistic causation: the manipulability theories of causation (MTPC) and the spacetime theories of probabilistic causation (STPC). Again, we do not consider counterfactual theories of causation because Butterfield has already satisfactorily shown that (¬PI-CAUS)* and hence (¬OI-HOL)* do not hold within such theories. So, we consider the two remaining strong candidates for theories of probabilistic causation. The upshot of our analysis will be the following.

Within MTPC, we show that a rigorous argument can be made in favor of (¬PI-CAUS)*. That said, MTPC do not allow for inferences about causation at the ontological level but only at the empirical level. Hence, (¬PI-CAUS)* is only supported in a weakened version: failure of PI* alone are indicative of empirically observable causal link between the parameters and the outcomes, whereas failure of OI* is not indicative of any causal link between the outcomes that has any empirical manifestation. Also due to the empirical leanings of MTPC, the issue whether or not (¬OI-HOL)* holds remains open within MTPC.

Within STPC now, we show that the mainstream interpretation fails in general. Either both failure of PI* and failure of OI* are indicative of a causal link, or none are. Thus, no STPC support the mainstream interpretation understood as in the program of "experimental metaphysics". That said, we shall give some plausibility arguments to the effect that (¬PI-CAUS)* and (¬OI-HOL)* could be supported in a weakened version, i.e. a version in which no inferences about the causal structure of the quantum domain are made, but only about about how we can best understand it.

The upshot of our analysis is that:

1. No theories of causation supports (¬PI-CAUS)* and (¬OI-HOL)* under the strong

interpretation suggested by the program of experimental metaphysics;

2. An empirical version of (¬PI-CAUS)* is well supported within MTPC;

3. An epistemological version of (¬OI-HOL)* is plausible in some refined STPC.

Thanks to our analysis, the conditions under which a mainstream-like interpretation holds will have been specified. Also, the mainstream interpretation as interpreted as a form of experimental metaphysics will be undermined.

Before we turn to the proper assessment of (¬PI-CAUS)* and (¬OI-HOL)* within precise framework of locality and causation, we shall consider the work by Arthur Fine, who has been challenging the mainstream interpretation since its inception. In Chapter 5, we distinguish between three strategies that Fine uses in attempting to undermine the mainstream interpretation. We show that none of these strategies succeed. The upshot will then be that Fine fails to prove that Local Realism is not at stake in the violation of the BI by quantum phenomena.

Chapter 5

Fine ways to fail to secure local realism

For more that twenty five years, Fine has been challenging the traditional interpretation of the violations of Bell inequalities (BI) by experiment. A natural interpretation of Fine's theorem is that it provides us with an alternative set of assumptions on which to put the blame for the failure of the BI to hold, and a new interpretation of the violation of the BI by experiment should follow. This is not, however, how Fine interprets his theorem. Indeed, Fine claims that his result undermines other interpretations, including the traditional interpretation in terms of local realism. The aim of this chapter is to understand and to assess Fine's claims. We distinguish three different strategies that Fine uses in order to support his interpretation of his result. We show that none of these strategies is successful. Fine fails to prove that local realism is not at stake in the violation of the BI by quantum phenomena.

5.1 Introduction: Bell-type theorems, Fine's theorem and its interpretations

5.1.1 Bell-type results, and hidden variable programs

A Bell-type experiment generally involves the following elements:

- A measurement set up with two separated parts, or "wings", A and B, and a source;
- On each wing, a measurement device (detector plus analyzer) which can be set to measure one of several observables on each side, labeled respectively A_1, A_2, $\ldots$, A_n and B_1, B_2, $\ldots$, B_n, with the A_i, B_j commuting for all (i, j) and the A_i, A_j and B_i, B_j non commuting for every $i \neq j$;
- A system, produced at the source, such that it can be measured at the two wings of the experiment. Traditionally, the system is understood as constituted by two parts which travel in opposite directions corresponding to the two wings.

One then considers the single and double outcome distributions $P(A_i)$, $P(B_j)$, and $P(A_i, B_j)$, usually for $i, j \in \{1, 2, 3\}$.

These distributions can be easily computed with the quantum algorithm. This algorithm does not allow one to specify which particular outcome is to obtain for a given setting of the experimental device; it only assigns probabilities over a spectrum of possible outcomes. Some experiments can be conducted which arguably correspond to measuring the outcome distributions. The experiments to date are almost uncontroversially[1] taken to be in agreement with the predictions of quantum mechanics.

A striking feature of Bell-type situations is the correlations obtaining between the outcomes of the two wings of the experiment. According to the quantum algorithm, there are

[1] There remains some debate on the so-called "detection loophole". For an analysis of the detection loophole, see the work of J.A. Larsson, for instance [78].

quantum states and experimental configurations such that, whereas we are in a state of complete ignorance about what outcomes are to obtain at each wing taken separately, we can be in a state of complete certainty on what outcome is to obtain in one wing once we know which outcome obtained on the other wing.

In general, there are two natural ways to interpret correlations: either they arise from an interaction between the two wings of the experiment or through a common cause. In our case, a last minute interaction would take place between the system that has been measured and the one that has not such that the results of the measurements would be perfectly correlated. However, actual experiments are generally taken as having the two wings so positioned that a last minute interaction would have to occur faster than the speed of light. This in turn is usually taken to conflict with Relativity Theory. So, explaining the correlations by a last minute interaction is usually taken as entailing a form of non-locality.

An alternative interpretation of the correlations is that the values of the outcomes were in fact determined before measurement and only revealed by measurement. In this case, the correlations stem from a common cause, i.e. the interaction of the two subsystems at the source. Probabilities are just representations of our ignorance of the real underlying situation on this view.[2] They are grounded in determinate properties, possibly hidden in the sense that we do not have any control of them.

It is the aim of hidden variable (h.v.) theories to assess which types of non-local probabilistic models can return all statistical predictions of quantum mechanics. In particular, hidden variable research programs seek to assess whether a model, in which probabilities are interpreted in terms of our ignorance of the underlying determinate properties, is possible.

It might seem very difficult to assess in general what types of probabilistic models can be constructed consistently with all statistical predictions of quantum mechanics. Bell's

[2]Throughout this chapter, we shall use the phrases 'ignorance interpretation of probabilities' and 'probabilities interpreted in terms of ignorance' to refer to this situation, that is, the situation in which probabilities derive merely from our ignorance, and are thus epistemic.

theorem[3] is a major achievement in this matter. Bell has managed to show that any local, either deterministic or probabilistic, determinate-value model[4] obeys numerical inequalities, i.e. the so-called Bell Inequalities (henceforth, the BI).

It so happens that, for Bell-type situations described above, statistical predictions of quantum mechanics violate these inequalities. The experiments which have been conducted[5] are usually taken to favor quantum mechanical predictions. By a simple modus tollens, the upshot is that no local realist model can give an account of all quantum phenomena. Fine's 1982 theorem provides an alternative derivation of the BI thereby making additional interpretations of the violation of the BI by experiment possible.

5.1.2 Fine's theorem and its possible interpretations

Fine proved the following theorem:[6]

Theorem 2 *The following conditions on the correlations of a Bell-type experiment with* $i, j \in \{1, 2\}$ *are equivalent:*

[3]The original proof of Bell Theorem, within a deterministic context, is in [12]. The result has been generalized to the stochastic case by Bell in [11] and by others, including [29],[28], [6], and [85]. The literature on Bell-type theorems is plentiful. Bell's original papers are an indispensable reference. Most of them are reprinted in the collection [14], with some others in the second edition [16]. A recent synthesis of the discussions concerning Bell's theorems can be found in [104].

[4]The issue of the precise definitions of Locality and Determinateness cannot be addressed here in any detail for it would take us off course. As explained above, the assumption of locality is a constraint which is taken to forbid superluminal interactions between the two subsystems. Determinateness, or, as is traditionally referred to, realism, is satisfied if and only if the measurement outcomes are determinate, namely each experiment will have definite results, which may be fixed deterministically or only stochastically by the complete state of the system, possibly together with the state of the measurement context before the measurement. In this chapter, we shall refer at the traditional derivation of the BI as the derivation in terms of *local realism* for convenience, despite the variety and ambiguity of what these terms refer to.

[5]The most well-known are the experiments conducted in Orsay by Aspect and his team: [8], [6] and, for a more popular presentation, [7]. Michael Redhead provides a table of the experimental results concerning BI up to the 90's ([93, p.108]). Many others have been conducted since, among which the ones by Zeilinger's team in Vienna and Gisin's team in Geneva are well known (see for example [134], and for a less technical presentation, [141]).

[6]The original proof is to be found in [48]; [49] is more philosophically concerned, and more ambitious in terms of the consequences Fine wants to draw from his result, and [52] is making the point in a non-technical way.

1. The BI hold for the probability distributions of the experiment;

2. There is a deterministic hidden variable model of the experiment returning the observed outcome distributions (singles and doubles);

3. There is a factorisable model of the experiment returning the observed outcome distributions (singles and doubles);

4. There is a joint distribution for all four observables of the experiment $P(A_1, B_1, A_2, B_2)$, compatible with the observed outcome distributions (singles and doubles) as marginals;

5. There are well-defined joint distributions for all pairs and triples of commuting and noncommuting observables compatible with the observed outcome distributions (singles and doubles) as marginals.

The result is impressive and multifold but two striking features deserve discussion. First, Fine's theorem provides us with alternative derivations of the BI in terms of the definition of certain joint probability distributions. He indeed shows that it is sufficient, for a probabilistic model to satisfy the BI, to have well defined joint probability distributions for all pairs and triples of observables, or to have well defined joint distributions for all four observables, whether these observables are compatible or not – let us call these PJPD, for Problematic Joint Probability Distributions. (We say"Problematic" because they cannot be correct because of the violation of the BI by experiment). Second, Fine's theorem establishes the converse of the original Bell derivation of the inequalities[7].

A natural interpretation of Fine's result is as follows. An alternative derivation of the BI is provided: the BI hold in any probabilistic model in which the PJPD are well defined. Alternative assumptions are uncovered on which we can put the blame in a Bell-type modus tollens argument. No model in which the PJPD are well defined can give an account of

[7] In this chapter we confine ourselves to the deterministic models of the BI, so we will not deal with case 3. Nonetheless, our conclusions apply in the stochastic case as well

all quantum phenomena. The ontological significance of this fact can be the subject of philosophical inquiry.

This is not the way in which Fine interprets his theorem. From 1982 on, he has been holding that his theorem does more than provide an alternative derivation from which additional interpretations of the ontological situation underlying Bell-type experiments can be made. Rather, Fine contends that the interpretation in terms of well defined joint probability distributions supersedes other interpretations of the violation of the BI. In particular, the traditional interpretation in terms of local realism would be threatened. Consider the three following quotations, two from the original papers, the other from a more recent one:

> Our investigations suggest that what the different hidden variables programs have in common, and the common source of their difficulties, is the provision of joint distributions in those cases where quantum mechanics denies them.[8]

> Finally, I believe that Proposition (1) – conjoined with the other two [items 2, 3, and 5 above]– shows what hidden variables and the Bell inequalities are all about; namely, imposing requirements to make well defined precisely those probability distributions for non commuting observables whose rejection is the very essence of quantum mechanics.[9]

> The Bell inequalities have a purely probabilistic content [...]. [10]

As it stands, such claim does not seem to hold. It is a simple matter of logic that the theorem does not undermine the other derivations, nor conclusions drawn from them. Further argumentation is needed to support Fine's strong interpretation of his results.

[8] [49, p.1309]
[9] [48, p.294]
[10] [53, p.3]

5.1.3 Fine's three lines of argumentation

Fine uses three different strategies to defend his strong interpretation. In this chapter, we shall object to each of these lines of argument in turn.

A first strategy Fine adopts to argue for his strong interpretation is to claim that his theorem uncovers an *additional assumption* in the traditional derivation, thus making an *alternative option for the modus tollens* available, an option in which the BI would be violated while locality and determinateness secured. According to this strategy, Fine wants his theorem to show that local determinate h.v. models include a hidden assumption which is sufficient to make them satisfy the BI.

However, as we shall point out in Section 5.2, Fine's argument only holds for a very restricted class of h.v. models. In particular, Fine never considers contextual h.v. models. For contextual h.v. research programs, the interpretation in terms of local realism is not superseded by Fine's interpretation of the violation of the Bell inequalities. Bell-type theorems and Bell-type experimental results are highly relevant to those research programs: indeed, it proves that they can be compatible with all quantum statistical predictions only if they are non-local.

Section 5.3 will focus on a second strategy of argumentation, which is based on a de facto argument. Fine has produced some local realist models of Bell-type experiments, the so-called Prism Models, which reproduce the violation of the BI. Prism Models are local deterministic classical probabilistic models returning the probability distributions of the actual Bell-type experiments. Fine seems to hope to conclude from the existence of Prism Models that local realism is not at stake in the violation of the BI by quantum phenomena: there are, de facto, some probabilistic models that violate the BI and in which local realism is secured.

Against this second line of argumentation, we shall point out that Prism Models are not models of quantum probabilities, but only of the actual experiments performed so far.

Prism Models are successful in returning the outcome distributions of Bell-type experiments because of a specific feature of the actual realization of the experiments: detector deficiency. However, Prism Models do not constitute a local realist model compatible with all statistical predictions of quantum mechanics. Prism models indeed contradict quantum theory in denying that a perfect experiment is in principle possible. Contra quantum theory then, they predict that the detector efficiency has a physical limit. Hence, their existence does not threaten the traditional view that no local realist classical probabilistic model can return all statistical predictions of quantum mechanics.

The last line of argumentation in favor of the strong interpretation of Fine's theorem is based on the fact that the theorem does not only show that the BI can be derived from the assumption that some joint probabilities are well-defined. It also shows the *equivalence* between these joint probabilities being well defined and the other sets of assumptions used in alternative derivations. Fine seems to claim that the converse of his new derivation, that is, that the BI's holding implies the PJPD to be well defined, supports his strong claim.

Fine interprets the converse of his new derivation in the following way: whatever its set of assumptions, any framework from which the BI are derivable requires, in Fine's terms, the "existence of well defined joint probabilities", especially for non-commuting observables. Now, Fine's point seems to be that the definition of these joint probabilities is just what quantum theory not only does not provide, but also strictly forbids. From this, Fine wants to deduce that h.v. models for quantum probabilities start right from the beginning with an inconsistent set of hypotheses.

The argument is not all clear. Svetlichny, Redhead, Brown and Butterfield have investigated this argument in [118]. They reconstruct Fine's argument as coming in two steps, each of which, we believe, involves controversial claims:

1. In accepting QM, the h.v. investigator is committed to rejecting the existence of joint probabilities for incompatible observables;

2. By the converse of the new derivation, any model from which the BI are derivable is committed to the existence of the PJPD, including joint probabilities for incompatible observables, whatever the other assumptions might be[11].

Svetlichny, Redhead, Brown and Butterfield have shown in [118] that 2. is not well supported. In Section 5.4 of the chapter, we shall recall their argument, and proceed to address 1. We shall point to an assumption that Fine seems to make in most of his arguments, namely, the "reality" of probabilities. Our objection will be rather simple: the acceptance of a theory does not include a specific interpretation of probabilities. Applied to 1., this means that, in accepting quantum mechanics, one is *not* committed to anything like "the existence" or the "non-existence" of well defined probabilities in the sense needed by Fine.

5.2 A hidden assumption in the traditional derivation?

One way to construct hidden variable models is to construe observables as random variables defined over a common classical probability space. Such construction is usually referred to as an "ensemble representation". It is a formal consequence of an ensemble representation that the joint distributions are well-defined for all pairs of observables, commuting or not. Further, these joint distributions are compatible with the singles as marginals. Hence, by Fine's theorem, some BI hold for such models. Fine holds that this supports the claim that his theorem uncovers a hidden assumption in the traditional probabilistic models, an assumption which alone is sufficient to make the BI to hold:

[11] Note that in their paper, Svetlichny, Redhead, Brown and Butterfield actually consider the contrapositive of 2:

2. By rejecting the joint probabilities to exist, the local realist is committed, by the converse of Fine's derivation, to expect the outcomes's distribution to violate the BI, whatever other characteristics of the models considered.

The version we consider is of course logically equivalent. It stresses a different aspect of Fine's criticism of the hidden variable research program: namely, not that the violation of the BI is directly given, but that the local realist starts off with an inconsistent set of premises. Fine uses both versions of the argument.

> [...] any ensemble representation must make well defined joint distributions for incompatible observables (like position and linear momentum, or spin in skew directions), for the function associated with these observables will be random variables over a common space and these always have well-defined joints.[12]

In this section, we shall first explain how ensemble representations indeed include the definition of the PJPD. We shall then proceed to criticize Fine's strong conclusion in pointing out that a large class of hidden variable theories do *not* construe quantum observables as random variables over a common classical probability space. For this class of h.v. theories, Fine's argument does not stand.

5.2.1 Ensemble representations and joint probabilities

A classical way to deal with statistical distributions of outcomes is to reduce them to determinate but unknown states. In our case, let the "complete state" of a system be the quantum state plus the set of hidden variables which together determine the results of measurements. This more complete specification of the state is denoted by means of variables λ. Whether λ is a single variable or a set is irrelevant here. The only relevant feature of the parameter λ is that it provides a complete description of the quantum system considered.

There is more structure needed for the construction of a determinate model for the given distributions on the basis of the consideration of the variables λ. Consider the set Λ of all possible hidden variables λ. Even if all we know about a system is confined to the probability distribution over a spectrum of possible outcomes given by $|\psi\rangle$, we want the system under consideration to be really in one of the complete states corresponding to some variables λ (or the state $|\psi\rangle$ really describes an ensemble of systems, each of which is in one of the complete states corresponding to the λ). For each $|\psi\rangle$, this set Λ is then structured as a classical probability space, that is to say, is equipped with a probability measure on its Borel sets $\mathcal{B}_\Lambda$.

[12] [51, p.45]

Thus, for a given $|\psi\rangle$, there is a probability density $\rho^{|\psi\rangle}(\lambda)$ of the possible hidden parameter λ over the space Λ. That is to say,

$$P^{|\psi\rangle}(I) = \rho^{|\psi\rangle}(\lambda)\, d\lambda \tag{5.1}$$

represents the probability that the hidden variables lie in the interval $I = [\lambda, \lambda + d\lambda]$.

What remains to be done is to provide a formal way of connecting the statistical distributions of outcomes with the probability measure over the variables λ. To this aim, every observable Q_i, $i \in \{1, \ldots, n\}$, is defined as a random variable on the common probability space Λ. In other words, every observable Q_i is represented by a real-valued function $[Q_i]$ defined over the domain Λ of the possible λ. For a given state, any observable has a range of determinate values associated with the possible λ. A consequence of construing observables as random variables over Λ is that the probability structure $(\Lambda, \mathcal{B}_\Lambda, \rho^{|\psi\rangle}(\lambda))$ induces one on the Borel sets $\mathcal{B}_{E_i}$ of the ranges E_i of the functions $[Q_i]$ corresponding to the observables Q_i (in general the E_i are some subset of the real numbers). A new probability measure is thus defined on the subsets F of the $\mathcal{B}_{E_i}$, that is, the probability $P^{|\psi\rangle}_{[Q_i]}(F)$ that the function $[Q_i]$ of the observable Q_i takes its value in F. It is defined in the following way:

$$\forall Q_i,\, \forall F \subset \mathcal{B}_{E_i}, P^{|\psi\rangle}_{[Q_i]}(F) = P^{|\psi\rangle}[([Q_i])^{-1}(F)] \tag{5.2}$$

So, the probability structure on the set of hidden variables is projected on the space of the possible values of each observable. Thus, the statistical distributions of outcomes given by $|\psi\rangle$ for a given observable are reduced to statistical distributions of hidden variables.

Further, since the functions $[Q_i]$ are defined as random variables on a common probability space, any Borel function $f([Q_1], [Q_2], \ldots, [Q_n])$ is in turn a random variable on the Cartesian product Λ^n, and there is a new probability measure defined on the Borel sets of its range, fully characterized by $\rho^{|\psi\rangle}(\lambda)$ on $\mathcal{B}_\Lambda$. In particular, for all pairs of observables $(Q_1,\, Q_2)$, whether compatible or not according to quantum formalism, the joint probability $P^{|\psi\rangle}_{[Q_1],[Q_2]}$

and the random variable $f([Q_1],[Q_2])$ is defined for all Borel function f such that they satisfy, for F_1,F_2, Borel subsets of respectively $\mathcal{B}_{E_1}$ and $\mathcal{B}_{E_2}$.

$$P^{|\psi\rangle}\left[([Q_1])^{-1}(F_1)\cap([Q_2])^{-1}(F_2)\right]=P^{|\psi\rangle}_{[Q_1],[Q_2]}(F_1\times F_2) \tag{5.3}$$

The probability that two observables take jointly some given values is fully determined by considering the intersection of the sets of hidden variables respectively corresponding to the given values.

Finally, the quantum expectation values are recovered as the expectation values of the functions corresponding to the observables along the the lines of the usual probability rules. That is, given a certain state, a simple integration over the whole range Λ of the possible λ provides the expectation value of any of the observables or of any Borel function $f([Q_1],...,[Q_n])$ of the observables.

$$< f([Q_1],...,[Q_n]) >^{|\psi\rangle}=\int_\Lambda f([Q_1],...,[Q_n])(\lambda).\rho^{|\psi\rangle}(\lambda).d\lambda \tag{5.4}$$

To summarize, such h.v. models amount to constructing an ensemble representation for all the observables, which includes the definition of joint probabilities for all observables returning the singles as marginals, whether these observables are commuting or not. By Fine's theorem, the Bell inequalities hold for any probabilistic model of this kind.

Granted then, it is the case that any probabilistic model in which the observables are construed as random variables over a common classical probabilistic space entails that the PJPD are well defined. For these models, whatever the set of other assumptions they may include, whether it be locality, realism or anything else, the PJPD being well defined alone can be blamed for their incompatibility with some of the quantum mechanical statistical distributions of outcomes. For these models, the violation of the BI do not need to involve anything else than a specific probability structure.

This alone, however, does not imply that Fine's theorem undermines the interpretation of Bell-type results in terms of local realism. For it is not the case that all hidden variable probabilistic models construe observables as random variables over a common classical probabilistic space. In particular, Fine never considers contextual h.v. research programs.

5.2.2 Hidden variable models outside the scope of Fine's argument

In arguing that his theorem shows that the traditional derivations of the BI include a hidden assumption about the definition of the PJPD, Fine seems to pursue two different aims. One aim is to dismiss h.v theories as a whole in saying that "what the different hidden variables programs have in common, and the common source of their difficulties"[13] is precisely the definition of the PJPD. Thus, h.v. research programs, whether local realist or not, would simply be dead ends. A second aim is to deflate the significance of Bell-type theorems and Bell-type results: the PJPD being well defined would be what the BI are all about. In particular, the violation of the BI by experiment would not mean that h.v. theories have to be non-local in order to be compatible with all quantum phenomena.

Fine's argument presented above fails to help to achieve any of these aims. This is because the classical construal of the observables as random variables over a common probability space, which Fine shows to include the assumption of well defined joint probabilities, is non-contextual. But, first, we did not need to wait for Fine's theorem to know that no non-contextual h.v. theory can model quantum probabilities. Second, there are contextual h.v. theories, which do not include the definition of the PJPD, and thus, which simply fall outside Fine's argument. Indeed, for such models, the derivation in terms of local realism remains highly relevant, for it forbids them to be local. Let us make these points in more detail[14].

[13] [48, p.1309]
[14] For similar points, see Shimony in [101].

Hidden variable theories are said to be contextual, if and only if the value of the observable obtained by any measurement process is not fully determined by the complete state of the system, $\lambda \in \Lambda$, but depends on what other quantities are simultaneously measured, or of the state of the measuring apparatus or any other details of the measurement context. Shimony 1984 made a distinction between algebraic and environmental contextuality. Algebraic contextuality is when the value of a quantity measured depends on other quantities measured. Environmental contextuality is when the value of a quantity depends not on other quantities measured but on other features of the measurement context.

The idea of contextuality is more familiar than it seems. Let us take a rather simple analogy. Consider a horse race. Whether Quick Silver, your favorite horse, is going to win the race or not may depend on various factors. One factor could be the degree of humidity of the ground: for example, Quick Silver, even though the best horse ever on a dry ground, might be very uncomfortable on a muddy soil. It seems also reasonable to think that whether Quicksilver is going to win or not depends on which other horses are running the race. If so, the outcome of an experiment in which a given quantity is measured (Quick Silver wins: 1, Quick Silver loses: 0) depends on both the environmental context (the state of the track) and on the algebraic context of the race (What other horses are "measured"). It is both environmental and algebraic contextual.

Now, the entire class of non-contextual h.v. theories were ruled out before Fine's theorem. Indeed, that there can be no non-contextual h.v. model compatible with all statistical predictions of quantum mechanics, was famously shown by Gleason in his profound theorem in 1957 for Hilbert space of dimension greater that two,[15] and in a simplified version of the theory by Bell in 1966,[16] as well as by Kochen and Specker 1967 theorem[17]. Fine's theorem

[15]in [57].
[16]in [13].
[17]in [75].

does not seem to provide anything new as far as non-contextual h.v. theories are concerned.[18]

Further, Fine cannot hope to dismiss all h.v. research programs on the basis that all non-contextual h.v. theories include the definition of the PJPD. The entire class of contextual h.v. theories simply falls outside Fine's argument. Bohm's theory is such a contextual h.v. theory, which returns all statistical predictions of quantum mechanics within the bounds of current empirical accuracy. It does not include the definition of the PJPD, and thus does not necessarily imply that the BI hold by Fine's theorem. This is because Bohm's theory does not use the classical construal of observables as random variables over a common probability space, but the framework of statistical variables.

A statistical variable is real-valued function with a probability measure directly defined on the Borel subsets of its range. This can be contrasted with random variables which we considered up to now. A random variable is a real valued function defined on the space of complete states, or on some subset of the space of complete states. In the case of a random variable, there is a probability measure on the possible values of the variable (which derives from the probability measure on the states). Therefore, a random variable is a statistical variable but the converse does not hold, as the example of Bohm's theory shows.

Within Bohm's theory, the result of a given measurement is encoded in the position of the "pointer", for which there is always a well defined probability distribution. While of course the result of a given measurement depends on the λ, which values of the λ are underlying which positions of the pointer depends on the context of measurement. Thus, even if the outcomes are construed as deterministically arising from an underlying physical situation, the observables are not construed as random variables on the common space Λ. Instead, the various experimental set-ups correspond to statistical variables.

Now, given two statistical variables, the joint distribution is generally not defined. By contrast, two random variables, as explained above, whenever they are defined on a common

[18]See Shimony in [101] for a discussion of the case of dimension 2.

domain, have their joint distributions well defined. Clearly then, by Fine's theorem, the BI do not necessary hold in a statistical variables model of the quantum distributions.

It is worth noting that the framework of statistical variables is precisely the framework that Fine has been advocating as the appropriate framework for quantum probabilities since (at least) 1968.[19] Fine argues convincingly that to construe the observables as statistical variables is sufficient to avoid the definition of the PJPD.

That said, Fine does not provide the framework of statistical variables with an h.v interpretation, that is, an interpretation in which probabilities only reflect our ignorance of the real physical situation. Rather, Fine claims that statistical variable models admits a natural interpretation he dubs "Minimal Realism".[20]

According to Minimal Realism, the following objects and properties are real:

1. physical objects – corresponding to the theoretical systems;
2. generic features which can take different forms – corresponding to the observables;
3. particular forms of the generic features – corresponding to the values of the observables;
4. distributions of probability on the spectrum of particular forms of the generic features – corresponding to the distributions of probability on the spectrum of values of the observables, ascribed by the quantum state.

Thus, the theoretical state of a quantum system is associated with an objective physical object (say, a quantum coin), an objective property (say, Quantum Face), a spectrum of objective forms of the property (Quantum Head and Quantum Tail), and the objective probability measure over this spectrum of forms (An even distribution for instance between Q-Heads and Q-Tails if our quantum coin is fair).

[19] See [45], [46], [47], and [51, chap. 4].
[20] See [51, p.160 sq.].

There are various aspects in which Minimal Realism is puzzling. That said, for this chapter, we shall insist only on the following point. In minimal realism, probability distributions are objective features of the world: they are second order properties of the object. This means that probabilities are properties of properties. Each property possesses as a property the probability distribution over its different forms. In taking the probabilities to be fundamental, Fine gives up the ignorance interpretation of quantum probabilities.

Fine never considers that an ignorance interpretation of the probabilities might still be possible. Such interpretation, however, is precisely what Bohmians advocate for the framework of statistical variables. Within Bohm's theory, quantities are interpreted either as categorical properties (for the quantities with continuous spectra such as position) or as dispositions (for all others). The dispositional properties considered by Bohm's theory can be seen as being reducible.[21]. Reducible dispositions are not ontologically significant: they reduce to categorical properties together with the experimental context. On Bohm's theory, measurement outcomes are determined by the underlying situation, including the initial positions of the particles, the initial wave-function, and the experimental context. The probability distribution associated with any given measurement arises from our ignorance of the underlying situation, in particular from our ignorance of the initial configuration of the particles.

So, even if the PJPD are not well-defined probabilities, it does not mean probabilities cannot receive an ignorance interpretation. H.v. models are possible despite Fine's theorem. Moreover, Bell has shown that the BI are derivable from local realism in a contextual setting. Thus the violation of the BI by experiment is relevant to contextual h.v. theories.[22] Indeed, it implies that, since they endorse the assumption of "realism", contextual h.v. theories have to be non-local in order to be compatible with all statistical predictions of quantum mechanics.

[21] See Clifton and Pagonis on this in [30].
[22] in [13].

Bohm's theory is one such non-local contextual h.v. theory.

Thus, Fine's first line of argument cannot deflate the significance of Bell-type theorems and Bell-type results to merely being that some joint probabilities are well defined.

5.3 A de facto argument: Fine's Prism Models

Fine is well known to have constructed local realistic models for the Bell experiments, the so-called Prism Models. Fine seems to believe that the existence of his Prism Models implies that the traditional interpretation in terms of Local Realism is threatened. Indeed, he argues that:

1. Prism Models are deterministic, local realistic models, where the joint probabilities are not defined and which return the single and double outcome distributions of actual Bell-type experiments;

2. Prism Models provide a probabilistic model of the quantum probabilities;

3. Hence: the traditional interpretation of Bell-type results in terms of local realism is threatened[23].

In the following, we shall present the construction of Prism Models in 1. and refute 2., so that 3. does not follow.

5.3.1 Prism Models

Prism Models are probabilistic models characterized by the special feature that observables do not in general have well defined probability distributions on all complete states

[23]See for example [50]. Szabo, in his [119], gives a very clear exposition of what Fine's Prism Models do and discusses possible future developments of them.

$\lambda \in \Lambda$. More precisely, there are some quantum states, for which there are some hidden variables, for which there are some observables, on the ranges of which no probability distribution is defined. From a mathematical point of view, the idea is just that, while observables are separately construed as random variables, they are defined on different subsets of Λ. Thus, observables are associated with partial real-valued functions on Λ. Using Fine's notation with a slight modification: a given observable Q_i is not defined on the entire space of complete states, but rather, given a quantum state, on a subset $\sigma(Q_i)$ of Λ.

It should be noted that this is different from saying that some observables have a zero probability associated with some range of outcome for a given state. In other words, Fine's construal of the observables as partial random variables does not entail that, associated with certain observables or experimental configurations, a given state has probability zero to display such or such outcome. Instead, in the framework of partial random variables observables are not in general defined on the entire space of hidden states. No probability distribution is defined for these states which fall outside the domain of the function of the observable considered.

Now, there is no reason to think that another observable Q_j will have its function $[Q_j]$ defined on the $\lambda \in \sigma(Q_i)$. In other words, $\sigma(Q_i)$ in general includes states in and outside the domain of the function $[Q_j]$ associated with the measurement of Q_j. If we happen to measure Q_i and Q_j at the same time, the set of hidden states for the joint systems on which the joint measurement is defined is the intersection $\sigma(Q_i) \cap \sigma(Q_j)$. So, whether you measure Q_i and Q_j together or separately makes a difference as to which set of complete states is selected. In some cases, the intersection may be empty for some observables, so that not all joint distributions are defined. Typically, in the case of a Bell-type experiment, the joint distribution over all observables is not defined, because the intersection of all domains of all the functions associated with all the observables is empty.

To complete the presentation of Prism Models, more needs to be said about their interpretation. That the observables are associated with partial real-valued functions on Λ is interpreted in terms of "*defectiveness*". For a given state $|\psi\rangle$, the λ encodes for determinate values of the outcomes, but it also encodes for some physical, predetermined property (or properties), "defectiveness", that makes the system under consideration suitable for certain measurements, and not for others. An Q_i-defective system just cannot respond to the experimental configuration (an analyzer-detector assembly) corresponding of the observable Q_i and hence will not "show up" in the sense that it will not be counted by the detector. Applied to a Bell-type experiment, given the experimental configuration or observable Q_i^A at the A end of the experiment, λ is in $\sigma(Q_i^A)$ if and only if an A-particle of type λ will be detected by any detector associated with a Q_i^A-analyzer.

With this construction in hand, Fine shows that there are some Prism Models for the actual Bell experiments with two observables on each end. That is to say, it is possible to divide Λ in subsets and define the four observables considered in the experiments over these subsets, such that the single and double probability distributions given by the model are consistent with the observed outcome distributions which violate the Bell inequalities.

Because they construe each observable as a random variable defined on a classical probabilistic state, Prism Models are models in which the probabilities are interpreted in terms of ignorance. That is to say, some underlying (even if possibly hidden) properties, associated with the variables λ, fully determine the outcomes of our experiments. So, Prism Models are determinate. Prism Models are also deterministic. Finally, Prism Models are local. There is no need for any superluminal information transfer to recover the correlations. Granted then, Fine has constructed local realistic models returning the statistical distributions of the actual experiments for dimension 2×2. This is an impressive result. Szabo, in his nice presentation of Fine's model, further proves the possibility of Prism Models for $n \times n$ dimensional systems.[24] This, however, does not suffice to support Fine's strong claim about the irrelevance

[24] in [119].

of the notions of locality and realism for the interpretation of Bell-type experiments. To put it bluntly, this is because Prism Models, while compatible with the actual experiments, are incompatible with quantum mechanics.

5.3.2 Prism Models, the experiments, and quantum mechanics

Fine's Prism Models are successful because of a specific feature of the *actual* realizations of Bell-type experiments, namely the deficiency of the detectors used. Our detectors just fail to catch all particles we send to them. Our statistical measure of Bell-type correlations thus takes into account only these particles that are "coincidentally" detected by both detectors. The usual way around the deficiency problem is to assume that the sample of particles we actually detect is a fair one. This is called the "enhancement hypothesis".[25] Fine's models draw on this feature to recover the probability distributions of the outcomes, interpreting what is usually taken as the deficiency of the detectors as to be reflecting a physical property of the systems, i.e. defectiveness.

While giving room for a local deterministic account of the actual experimental results, the construal of the observables as *partial* functions on Λ makes Prism Models stand in contradiction to quantum mechanics (as normally construed). At least if interpreted in terms of defectiveness, Prism Models predict that not all particles will be detected by a given experimental context. In other words, Prism Models predict that there is an irreducible deficiency in the detectors. Even if not interpreted in terms of defectiveness, however, they fail to make predictions (even probabilistic ones) for some runs of the experiments, which quantum mechanics instead makes.

On the contrary, quantum mechanics predicts that any quantum system should respond to any given experimental question. Indeed, according to the usual quantum measurement theory, any quantum system can be appropriately coupled to some other system which will

[25] On this, again, see [78] and reference therein.

act as a perfect measuring apparatus.[26] Any deficiency in the experiment is in one of our detectors – it is a technical problem – and does not reflect any property of the system being measured.

Because Prism Models contradict quantum mechanics, their existence does not support the strong claim held by Fine, that there are models consistent with all quantum statistical predictions in which local realism is secured. It also does not support the claim that locality and realism are not at stake in the violation of the BI by quantum phenomena. Fine suggests that it is *because* the joint distributions are not defined that Prism Models can violate the Bell Inequalities in the right way. He claims that prism models keep the structural features of the probabilistic model which codify the "local realistic" character of the observables, while avoiding the assumption of the existence of the joint probabilities, which he deems responsible for the Bell inequalities to obtain. We have argued that the real reason why Prism Models are successful is our actual detectors' deficiency. Further, Fine's way to avoid that the PJPD makes Prism Models stand in contradiction with quantum theory. While providing an alternative model for the actual statistical results, Prism Models do not provide an alternative model for quantum probabilities.

That said, that Prism Models are local deterministic models of the experiments to date, is, we believe, worth considering for further investigation. It would be certainly worth pursuing a research program to assess which one of the two models the experiments favor, that is, to experimentally assess whether there is a detector efficiency limit. This probably deserves more attention, especially if one considers the importance of recent positive experimental developments in quantum physics using Bell-type situations and entangled systems in general (in quantum cryptography for example). It would be interesting to assess whether Prism Models can give an account of these experimental protocols (perhaps with suitable provisos

[26]We do not consider the limitations shown by Araki and Yanase in [1]. They can be disregarded for present purposes.

related to detector efficiency). Szabo shows how Prism Models are physically motivated by the actual realizations of the experiments.[27] He further investigates to what extent Prism Models can be refuted by experiment on the basis of the detection efficiency rates. His conclusion is that actual experimental results still do not constitute a serious objection to Prism Models. Fine's models might be more relevant than they seem.

If further experimental investigation was favorable to Prism Models, some further interpretational work would still have to be done. The interpretation in terms of defectiveness seems unsatisfactory as it stands. What we have is that there are some complete states for which some observables cannot be measured. Fine does not give any further interpretation of this property. At least three further interpretations are possible. First, if an observable corresponds to some determinate property-assignment, one interpretation is that the property corresponding to the observable is not measurable. Unless Fine could provide an explanation for why this happens, such an interpretation goes against one of the most fundamental assumptions in the methodology of experiment about the possibility of measurement. Moreover, it might run into problems with Kochen-Specker-type theorems. Another interpretation would be that the observable corresponds to a generic property, but the value of this property is unmeasurable because it is indefinite. The idea is that some properties do not have any determinate value, that could be correlated either deterministically or stochastically with the measurement context. This could correspond to the notions of "spread quantities" or "inexact values" that have been defended by some authors.[28] However, Fine seems to reject this option himself: he accepts the objections leveled against such notion of spread quantities.[29] Yet another interpretation is that the observable corresponds to a generic property, but there are some states that simply do not "have" this generic property, in any form, exact or spread. This obviously needs more fleshing out, but it might be what

[27]in [119].

[28]See for example Teller in [121], [122], and [124].

[29]See [51, p.161, note 16].

Fine has in mind, considering the concluding remark of [49]:

> After all, if we hold that probabilities (including joint probabilities) are real properties, then some observable may simply not have them.[30]

That said, Fine never seriously investigates such issues. Indeed, Fine himself admits that Prism Models are somewhat too "cheap" and too "easy".[31] The aim of constructing these models was primarily negative, that is, it was to undermine the traditional interpretation in terms of local realism. This, however, as we have argued, is just not what Prism Models can be used for.

5.4 Beyond the alternative? Fine's strong claim

Fine's last line of argumentation in favor of his strong claim uses the *converse* of his derivation of the BI, that is, that any model in which the BI hold is also a model in which the PJPD are well-defined. Fine's theorem indeed gives the equivalence between the BI's holding and the joint probabilities being well defined. According to Fine, this means that whatever the set of assumptions, any framework from which the BI are derivable includes the assumption of, in Fine's terms, the "existence of well defined joint probabilities", especially for non-commuting observables. Now, Fine's point is that the definition of these joint probabilities is just what quantum theory not only does not provide, but also strictly forbids. From this, Fine wants to deduce that hidden variable programs start right from the beginning with an inconsistent set of hypotheses. Equivalently, in accepting quantum theory, the hidden variable investigator is directly committed to the non existence of the joint probabilities for incompatible observables, and hence to the BI to be violated, which contradicts local realism or any other set of assumptions used.

[30] [49, p.1310]

[31] See [51, p. 56 sq.]. Fine formulates the same kind of qualifications as to the philosophical or physical significance for his local model in response to the Hardy theorem in [53].

Fine's argument can be reconstructed as coming in two steps:

1. In accepting QM, the local realist is committed to rejecting the existence of joint probabilities for incompatible observables;
2. By rejecting the existence of joint probabilities, the local realist is committed, by the converse of Fine's derivation, to expect the outcome distributions to violate the BI, whatever other characteristics of the model considered.

Svetlichny, Redhead, Brown and Butterfield have investigated this argument.[32] They accept 1. but show that 2. is not well supported. According to them, 2. would hold, and Fine's theorem would threaten the traditional interpretation in terms of local realism as well as the strength of other proofs, only if the BI's holding for the outcome distributions implied "physically real" joint distributions. The point is that the local realist might be happy to accept the joint probabilities as being defined formally, but without any ontological commitment. In the following, we shall briefly sum up their argument against 2. and then modestly contribute to undermining Fine's argument in raising some objections against 1.

Let us adopt the following notation:

- (Com): all observables pairwise commute
- (BI): some Bell-type inequalities hold
- (R_{jd}): the joint distributions are physically real.

Svetlichny et al. convincingly argue that what Fine shows is that there is a condition (jd)[33], which is stronger than conditions 4 and 5 in Fine's theorem alone, such that:

$$(jd) \leftrightarrow (Com)$$

[32] in [118].
[33] (jd) stands for joint distributions.

$$(jd) \to (BI)$$

$$(jd) \to (R_{jd})$$

Now, step 2. which Fine needs in his argument translates as:

$$(BI) \to (R_{jd})$$

However, Fine fails to show:

- either directly that:

$$(BI) \to (R_{jd})$$

- or that

$$(BI) \to (jd)$$

By contrast, Svetlichny et al. show that:

- It is doubtful that the BI's holding is sufficient for real joint probabilities to exist, and
- there are cases where the BI hold and (jd) fails.

Hence, 2. is undermined. It is not true that all the frameworks from which some BI are derivable imply the existence of *real* joint probabilities which quantum theory forbids. Svetlichny et al. conclude that alternative proofs of Bell type theorem do avoid commitment to the existence of the joint distributions for non-commuting observables.

This is a good point against Fine's strong interpretation of his theorem. We believe that there is more to say though, in particular concerning 1. , i.e. that in accepting QM, the local realist is committed to rejecting the existence of joint probabilities for incompatible observables.

It seems to us that 1. consists in at least three components. The first component is the underlying assumption that hidden variable programs are committed to accepting quantum theory. The second component is that to accept quantum theory is sufficient for being realist about quantum theory, in particular about quantum probabilities. This is to say that to accept the theory commits one to a certain interpretation of the probabilities involved in the theory. The final component is that to be realist about a theory that contains probabilities is sufficient for being committed to the "existence" of only the probabilities that are well defined within the theory. Our contention is that none of these components is trivial. Each can be avoided in hidden variable programs, so that Fine's argument does not stand.

First, not all hidden variable programs presuppose the quantum formalism. As Shimony systematically points out, there are different types of h.v. programs, with different sets of assumptions and tackling different issues.[34] Now, an important part of the h.v. research programs aims at assessing whether some specific kind of probabilistic models are compatible with all statistical predictions of quantum mechanics. The construction of the probabilistic models generally does not presuppose the quantum formalism, quite the contrary. Famously, the original derivation of the BI by Bell does not presuppose the quantum formalism. The point of some of these research programs is to show what kinds of probabilistic models are *not* compatible with all statistical predictions of quantum mechanics. One can then investigate the significance of the failure of such models from an interpretational point of view: what the world could be like, since such and such probabilistic models are impossible? Such investigation guides other, more positive, research programs, which try out new theories which are empirically equivalent to standard quantum mechanics, up to our current level of empirical accuracy. Bohm's theory is obviously a case in point (although historically it developed in a more complicated manner).

Concerning the second component, it is not clear that Bell-type research programs are as committed to scientific realism as Fine wants them to be. To accept a theory can involve at least two different options:

[34]See [101].

1. to believe that the theory is true or approximately so, or,

2. to take the theory as empirically adequate.

One might want to take an empiricist stance towards quantum theory, but still investigate the interpretation of the violation of the BI by the experimental outcomes. Famously, van Fraassen takes such an empirical stance. However, he still contributed to the work on the interpretations of the Bell-type experiments. His interpretation is roughly that Bell-type experiments exhibit some phenomena which do not fit into one of our favorite models for scientific explanations: the common cause model. He argues that this is casting doubt on the legitimacy of the demand on our scientific theories that they give explanations in terms of common causes. We do not see how Fine could argue that an anti-realist like van Fraassen is committed, in investigating the derivation of the BI and their violation by experiment, to the existence of the joint probabilities in any ontologically robust sense.

From a more general point of view, we would like to stress that to accept a probabilistic physical theory does not involve any commitment to any particular interpretation of probabilities[35]. Scientific realists as well as scientific empiricists can choose to interpret the probabilities involved in the theory as being:

(a) either objective or subjective;

(b) either reducible or irreducible to physical properties.

That some probabilities can be defined formally does not imply that they "exist" in any ontologically strong sense.

Concerning the third component, i.e. that to be realist about a theory that contains probabilities is sufficient for being committed to the "existence" of only the probabilities

[35]It is striking that the core of many arguments of the objectors to Fine up to now has been to vary the interpretations of the probabilities in Fine's results: Shimony and Redhead consider counterfactual interpretations, Svetlichny et al. consider a Church-von Mises interpretation. The reason for this, it seems to us, is that Fine's argument hinges on a specific interpretation of the probabilities. Fine seems to implicitly assume that to accept a probabilistic physical theory commits one to one particular interpretation of the probabilities.

that are well defined within the theory, we maintain that it is as problematic as the first two, because the converse of this last statement holds as well: that some probabilities cannot be defined formally does not imply that they do not "exist". The point is easy to make in the case where one takes an empirical stance toward quantum mechanics. In this case, that some probabilities are not formally defined in the theory does not imply that they do not "exist", since a theory that pretends only to be empirically adequate does not preclude the "existence" of probabilities not contained in the formalism itself. But this holds for the non empiricists too. That some mathematical objects are not defined in a given formalism does not imply anything at the ontological level, even if one is a realist about the quantum formalism. For example, that deterministic trajectories for the particles cannot be defined in the standard formalism of quantum mechanics does not imply that they cannot exist at the ontological level. A realist Bohmian would obviously believe that they exist, while possibly being just as realist about the quantum mechanical wave functions.

Our point is that the ontological status of the counterparts of mathematical objects in the world has to be strictly distinguished from the formal construction of these mathematical objects in a given formalism. To distinguish, within the formalism, what is physically significant from what is an artifact of the mathematical construction, is, we believe, precisely one of the most important aims of the interpretational work on physical theories. The upshot is that, in 1. as well as in 2. , the main difficulty in Fine's argument is the confusion between the formal definition and the objective "existence" of probabilities.

5.5 Conclusion

We have argued that:

- Fine's argument that the definition of joint probabilities is a hidden assumption in the traditional derivations of the BI holds only for a restricted class of h.v. theories, which were ruled out by previous theorems;

- Fine's argument, to the effect that the existence of his Prism Models is a de facto argument against the traditional interpretation of Bell-type results, does not hold, because Prism Models are incompatible with some quantum statistical predictions;

- Fine's argument that the converse of his derivation shows that h.v. research programs start off with an inconsistent set of assumptions holds only under strong assumptions about the ontological status of probabilities.

It seems to us, then, that Fine fails to prove that his theorem has to be interpreted in a strong way, namely as undermining the traditional interpretation of Bell-type results in terms of local realism. Bell-type theorems and Bell-type results are still relevant for the interpretation of quantum theory and quantum phenomena.

That said, the framework of statistical variables, which Fine has been advocating as the appropriate one for quantum probabilities, is interesting, especially when fleshed out with a Bohmian interpretation. Further, Fine's Prism Models constitute a competitive model for the outcomes distributions to date, and should probably be given more attention.

Chapter 6

Locality in Bell-type phenomena

6.1 Introduction

Recall (PI-LOC)*:[1]

(PI-LOC)* : a violation of Parameter Independence* constitutes a case of non-locality while a violation of Outcome Independence* does not – non-locality being understood as in Special Theory of Relativity (STR).

In this chapter, we propose a spacetime framework for the probabilities considered in Bell-type situations, a rigorous definition of locality within such framework, as well as spacetime versions of Outcome and Parameter Independence*, so that (PI-LOC)* can be properly assessed. We show that, within the spacetime framework proposed, part of (PI-LOC)* is rigorously supported.[2]

Our results can be seen as an extension of John Earman's paper "Locality, non-locality and action at a distance: A skeptical review of some philosophical dogmas"[3]. In that paper,

[1]See Definition 8 in Subsection 4.2.2.

[2]Note that this result is not contradictory to Jones and Clifton's theorem presented in Chapter 4. What they show is that a violation of Outcome Independence is generally indicative of an underlying causal relationship. We shall argue that such causal influences do not amount to non-locality.

[3][41]

Earman argues that Factorizability is not a locality condition. We extend Earman's argument in two ways. First, Earman only investigates the issue of whether Factorizability is a locality condition; he does not address either Parameter or Outcome Independence*, on which we shall focus in this chapter. Second, Earman's discussion is mostly confined to the deterministic context: we generalize the discussion to the stochastic case.

In Section 6.2, we specify a spacetime structure and how Bell-type situations can be embedded in it. To this aim, we shall follow Earman and define locality as Einstein Locality. With this in hand, we shall propose (and argue for) our formulations for the spacetime versions of Factorizability*, OI* and PI* (STFAC*, STOI* and STPI*, respectively).

Section 6.3 deals with the issue whether or not Einstein Locality entails STPI* and/or STOI*. We begin with a discussion of how to construe Einstein Locality in the stochastic context. Bell has proposed a condition of locality in the stochastic case which he dubs "local causality".[4] We shall argue that Bell's condition is not an acceptable locality condition. We shall propose a natural generalization of Einstein Locality to the stochastic case, or Stochastic Einstein Locality (SEL).[5] Finally, we shall show that SEL implies STPI* but not STOI*. Hence, STPI* is a locality condition, at least in the sense that failure of STPI* is a case of non-locality,[6] while STOI* is not. Thus, most of (PI-LOC)* is supported.

Section 6.4 addresses what STOI* is. We argue that STOI*, when applied to Bell-type situations, is just a spacetime version of the PCC. In order to do so we discuss what an appropriate spacetime version of the PCC is. Our discussion relies on remarks in Earman's paper; however, the explicit formulation of the spacetime version of the PCC (STPCC) given is ours. We will give a rigorous argument to the effect that failure of STOI* is equivalent to

[4] [15, p. 54].

[5] For this, we shall partly rely on both Earman in [41] and Hellman in [64].

[6] The converse does not hold, that is: STPI*'s holding is not sufficient for SEL to be secured. So, strictly speaking, the mainstream interpretation does stand at least in its original guise. That said, the defendant of the traditional interpretation can content himself with STPI* being a necessary condition of locality: it is all he needs for a modus tollens in the interpretation of Bell-type experiments.

failure of the PCC. Depending on how one interprets failure of the PCC, the significance of failure of STOI* may change. We leave open the question of how to interpret failure of the PCC, except for one interpretation that we reject; namely, that failure of the PCC is failure of locality. We extend Earman's arguments to support this contention.

6.2 A spacetime framework for probabilistic conditions

6.2.1 A rigorous spacetime framework

Locality is a a space-time notion. Yet, violations of certain probabilistic conditions are supposed to be indicative of failure of locality. For such claims to make sense, the events for which certain probabilistic conditions hold will have to be embedded in spacetime. In order to provide a rigorous assessment of whether failure of Factorizability*, PI* and OI* threaten locality or not, what we need is a specification of how events are to be embedded within spacetime. Towards this end, we discuss locality in conjunction with a rigorous spacetime framework in this section.

A minimal spacetime framework should thus provide a notion of localization of events within the spacetime structure that is considered. In particular, if the notion of spatial separation and spatially separated systems are to be precisely defined, it should be possible to localize the systems under consideration within spacetime.

This requirement is difficult to fulfill in standard quantum mechanics (henceforth SQM), because SQM is not a spacetime theory. In effect, SQM provides no representation for quantum magnitudes as spacetime quantities. So, it is not obvious how to discuss locality issues rigorously.[7]

[7] That said, it should not be impossible either. One possibility is to endorse the Copenhagen interpretation in a version that takes the apparatuses to be classical. In this way, one can say that the apparatuses are located in classical spacetime. For our purposes this is an unattractive option. The project we are analyzing is experimental metaphysics whose conclusions should be independent of any interpretation of quantum mechanics.

We shall suppose that a spacetime perspective can be taken on quantum events. Within ordinary relativistic spacetime, quantum events are taken to be localizable in extended spacetime regions. We propose that an event e in region R can be represented by the total physical state over R. We will not consider any spacetime point-events in order to avoid singularities.[8] Whenever we shall refer to events embedded in spacetime, we shall specify the regions on which these events are localized. We denote event e located in region R by e_R.

What is at issue is whether certain probabilistic relations between events or failure thereof is indicative of non-local interactions between events. In order to address that issue we must examine what counts as a local interaction in relativistic spacetime.[9]

We first need the following notions: domain of influence, time slices, domains of dependence, and Cauchy surfaces. Let us begin with domains of influence. Let $\mathcal{M}$ stand for a spacetime. The domains of influence of a point p of $\mathcal{M}$ is defined as follows:

Definition 18 – ***Domain of influence***

The future (respectively past) domain of influence $I^+(p)$ ($I^-(p)$) of a point $p \in \mathcal{M}$ is the set of all the points q of $\mathcal{M}$ such that there exists a future(past)-directed timelike curve which begins at p and ends at q.

Figure 6.1 represents the past and future domains of influence of a region R of $\mathcal{M}$. One contrasts domains of influence with domains of dependence. In order to define domains of dependence, we need the notion of time slice:

Definition 19 – ***Time slice***

A time slice S of $\mathcal{M}$ is defined to be a spacelike surface $S \subset \mathcal{M}$ without edges.

With this in hand, one can define the past and future domains of dependence of a time slice S in the following way:

[8] Note that Reichenbach considers only point events, as opposed to Butterfield (Compare [95] and [21]).

[9] We restrict discussion to spacetimes in which Cauchy surfaces can be defined. See below for the definition of a Cauchy surface.

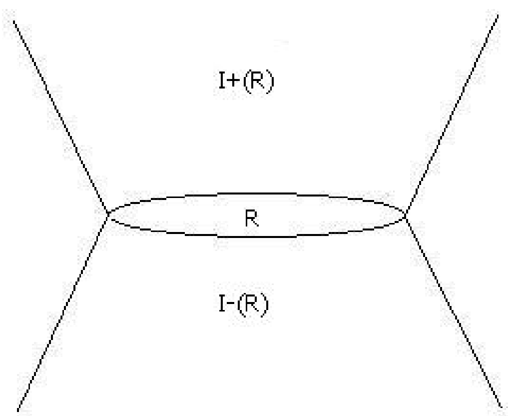

Figure 6.1: Domains of influence

Definition 20 ***Domains of dependence***

The future *(respectively* past*)* domain of dependence $D^+(S)$ $(D^-(S))$ *of a timeslice* $S \subset \mathcal{M}$ *consists of all those points* $p \in \mathcal{M}$ *such that every past (future) directed timelike curve that passes through* p *and has no past (future) endpoint meets* S*. The total domain of dependence of* S *is* $D(S) \equiv D^+(S) \cup D^-(S)$*.*

This is illustrated by Figure 6.2. With this in hand, we can finally introduce the notions of (future and past) Cauchy surface:

Definition 21 – ***Cauchy surface***

A timeslice S *is said to be a future (respectively past) Cauchy surface for* $\mathcal{M}, g$ *just in the case that all the points of* $\mathcal{M}$ *to the future (past) side of* S *are in* $D^+(S)$ $(D^-(S))$*.*

S *is a Cauchy surface simpliciter just in case it is both past and future Cauchy.*

The notion of Cauchy Surface is closely related to the notion of determinism, as well as to our favorite notion of locality: Einstein Locality. In short, Einstein Locality simply states

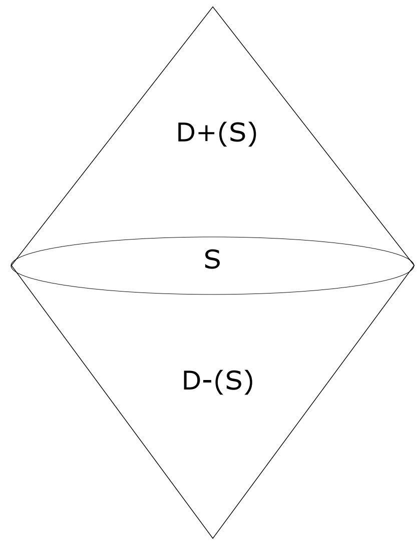

Figure 6.2: Domains of dependence

that an event localized within a certain region R of spacetime should be determined by an appropriate slice S across its backward light cone. But let us use Earman's definition of Einstein Locality.[10]

To do so, we need to introduce the notion of global Laplacian determinism:

Definition 22 – ***Laplacian Determinism***

If two models agree on appropriate data on a Cauchy surface then they agree everywhere.

It is also useful to introduce the notion of local Laplacian determinism:

Definition 23 – ***Local Laplacian Determinism***

Local Laplacian Determinism demands that the state on a region R be determined by the state on any spacelike surface S such that $S \subset I^{-}(R)$ and $R \subset D^{+}(S)$, and likewise with past and future interchanged.

This is illustrated by Figure 6.3.

[10]See [41, p.462]For another formulation, see [64].

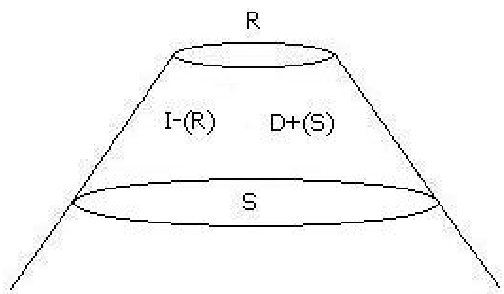

Figure 6.3: Laplacian Determinism: The state on R is determined by a time slice S such that S $S \subset I^{-}(R)$ and $R \subset D^{+}(S)$.

Finally, Einstein locality (EL) can be formulated:

Definition 24 – ***Einstein Locality*** – ***EL***

A theory satisfies Einstein Locality if when it is Laplacian deterministic it is locally so.

We now have the minimal spacetime structure needed for the connection between probability conditions and locality to be properly discussed.

6.2.2 Factorizability*, PI* and OI* in spacetime

We shall now propose formulations of Factorizability*, Outcome and Parameter Independence* within our minimal spacetime framework. To this aim, let us consider a Bell-type situation. What we are given are the outcome-events a and b (for variables A and B). In agreement with the above, we consider that a and b are located on two regions of spacetime R_1 and R_2. Now, in order to be able to assess whether the correlations between A and B are the result of a non-local underlying physical process, we need to make R_1 and R_2 spacelike separated.

The relativistic spacetime structure then gives us the past and future domains of influence, or backward and forward light cones of R_1 and R_2. Let S stand for the whole time slice across the backward light cones of R_1 and R_2, before they get separated, and such that $S \subset I^-(R_1 \cup R_2)$ and $(R_1 \cup R_2) \subset D^+(S)$. Let S_1, S_2 and S_3 be spacetime regions constituting a partition of S, so that S_1 (respectively S_2) correspond to the part of the slice within the past light cone of R_1 (respectively R_2) which is not within the past light cone of R_2 (respectively R_1), and S_3 corresponding to the part of the slice lying in the intersection, as in Figure 6.4.

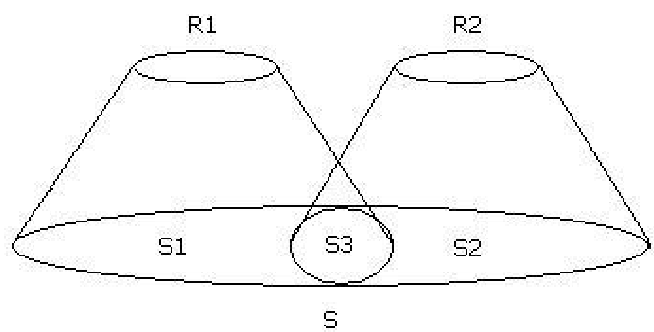

Figure 6.4: A Bell-type situation in spacetime: the outcome-events are located on regions R_1 and R_2, S is a time slice across the backward light cones of R_1 and R_2, such that $S \subset I^-(R_1 \cup R_2)$ and $(R_1 \cup R_2) \subset D^+(S)$.

We want to assess whether or not the various underlying causal paths that could be considered responsible for the observed correlations correspond to non local processes. In order to do so, we need to connect the spacetime regions described above with the various causal factors possibly influencing the phenomena. Remember that we considered the following potential factors: the complete state λ, the surface variables i and j, and the hidden variables γ and δ of the apparatuses. What spacetime regions can these factors be localized to?[11]

Let us consider first the time slice S. The total physical state σ over S includes the complete state λ. Now, the question of how λ is to be localized on S is a little tricky. λ certainly depends on some parts of S_3. That said, the state of S_3 and λ do not necessarily coincide. On the one hand, there is no a priori reason to consider that λ depends entirely on S_3. In particular, nothing in the ordinary relativistic spacetime framework implies that the complete state of the quantum systems involved in Bell-type situations be localized in the intersection of the backward light cones.[12] On the other hand, it is not clear for now that λ exhausts the state of S_3. The localization of the states of the apparatuses needs to be discussed in order to settle that matter.

The total physical state on S also includes the states of the apparatuses, surface and hidden. To localize these within our framework is again not easy, but a little more can be said about the states of the apparatuses than about the complete state λ of the system under consideration. We can legitimately localize the states (surface and hidden) of the two apparatuses on S_1 and S_2 respectively, so that they stay spacelike separated. It seems physically possible to make sure that the two apparatuses do not interact with each other.

What else is there on S? Arguably, if we consider that $\{\lambda, i, j, \gamma, \delta\}$ are the only potential causal factors relevant to Bell-type situations, then they should exhaust the total physical state on S relevant to Bell-type situations. Note however that there is no reason to hold that the regions of S_1 on which $\{i, \gamma\}$ and λ depend be mutually exclusive. Similarly for

[11] Strictly speaking this language is misleading. States and variables are mathematical objects and are not properly speaking located anywhere. What we mean by localizing states and variables is localizing the properties of systems on which the states and values of variables supervene. For brevity though, we will continue to use the language of localizing states and variables.

[12] We shall clarify this statement later on in Section 6.4 by explaining why Einstein locality does not imply that two events localized on spacelike separated regions should be determined by the state over a region of the intersection of their backward light cones.

$\{j, \delta\}$ and λ on S_2. In both cases, a tight localization of the system may be difficult to obtain – especially if the systems are field-like. Since we do not need such localization for our argument, we allow for some overlap. We only require that the regions on which $\{i, \gamma\}$ and $\{j, \delta\}$ depend do not overlap. Call the regions on which $\{i, \gamma\}$ and $\{j, \delta\}$ depend S_1^* and S_2^* respectively. The region on which λ depends may overlap with S_1^* and S_2^*. Call the region on which λ depends S_3^*, may extend outside of region S_3.

In the end then, we shall consider that $\{\lambda, i, j, \gamma, \delta\}$ are localized as in Figure 6.5.[13]

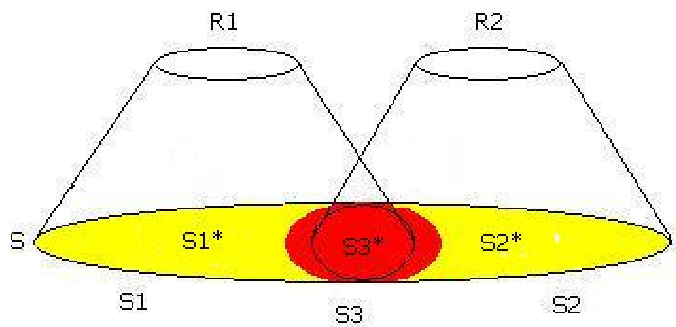

Figure 6.5: A Bell-type situation in spacetime: localization of the potential causal factors.

We now have a spacetime representation of Bell-type situations. With this in hand, we can formulate the spacetime versions of Outcome Independence* and Parameter Independence* (STOI* and STPI*).

Remember that OI* is the requirement that the outcome-events on R_1 and R_2 be statisti-

[13]One could object that our representation of Bell-situation in spacetime does not take into account whether $\{\lambda, i, j, \gamma, \delta\}$ will evolve, deterministically or stochastically, between S and R_1, R_2. We should then localize the causal factors on the extended regions stretching from S to just below R_1 and R_2. This is what Butterfield does in [21]. We believe we do not need to do so in the context of this chapter. The condition SEL, defined in Section 6.3.3, obviates this problem.

cally independent from each other, conditional on the hidden states of the system undergoing measurement and the measurement devices. Within our spacetime framework, this translates into: STOI* prescribes that the outcome-events a and b, localized on the two relatively spacelike regions R_1 and R_2, are statistically independent, conditional on the total physical state on an appropriate slice across the backward light cones. That is to say:

Definition 25 – ***Outcome Independence*: spacetime version – STOI****

A probabilistic model for Bell-type situations satisfies STOI if and only if for all* a_{R_1}, b_{R_2}, $i_{S_1^*}$, $j_{S_2^*}$, $\lambda_{S_3^*}$, $\gamma_{S_1^*}$ *and* $\delta_{S_2^*}$*:*

$$p(a_{R_1}, b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}).p(b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*})$$

On the other hand, Parameter Independence* requires that an outcome-event at one end be determined only by the apparatus parameters on its end, and not by the parameters on the other end. Translating within our framework, STPI* requires that the probability of an outcome-event a, localized on a region of spacetime R_1, be determined by the total physical state on an appropriate slice across the backward lightcone of R_1. That is to say:

Definition 26 – ***Parameter Independence: spacetime version – STPI****

A probabilistic model for Bell-type situations satisfies STPI if and only if for all* a_{R_1}, b_{R_2}, $i_{S_1^*}$, $j_{S_2^*}$, $\lambda_{S_3^*}$, $\gamma_{S_1^*}$, $\delta_{S_2^*}$, $i'_{S_1^*}$, $j'_{S_2^*}$, $\gamma'_{S_1^*}$ *and* $\delta'_{S_2^*}$*:*

$$p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j'_{S_2^*}, \gamma_{S_1^*}, \delta'_{S_2^*});$$
$$p(b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(b_{R_2}|\lambda_{S_3^*}, i'_{S_1^*}, j_{S_2^*}, \gamma'_{S_1^*}, \delta_{S_2^*}).$$

The spacetime version of Factorizability* (STFAC*) follows straightforwardly from the above:

Definition 27 – ***Factorizability*: spacetime version – STFAC****

A probabilistic model for Bell-type situations satisfies STFAC if and only if for all* a_{R_1}, b_{R_2}, $i_{S_1^*}$, $j_{S_2^*}$, $\lambda_{S_3^*}$, $\gamma_{S_1^*}$, *and* $\delta_{S_2^*}$*:*

$$p(a_{R_1}, b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, \gamma_{S_1^*}).p(b_{R_2}|\lambda_{S_3^*}, j_{S_2^*}, \delta_{S_2^*}).$$

It should be easy enough for the reader to figure out the spacetime versions of Local and Global Hidden Parameter Independences. Since we shall not need them, we skip them here.

We are thus provided with a rigorous spacetime framework, and with spacetime versions of the various probability conditions about which we want to assess whether or not they are locality conditions or not. We shall begin with STPI*. That said, before we can assess whether STPI* is a locality condition or not, we need to discuss the issue of how to construe Einstein Locality in the stochastic case.

6.3 STPI* and SEL

In this section, we shall argue that STPI* follows from a version of Einstein Locality in the stochastic context within our spacetime framework. We shall first discuss the issue of how to construe locality in the stochastic case. In Subsection 6.3.1, we present Bell's proposal for a formulation of locality in the stochastic case. We call Bell's condition SL, and argue that it is not a reasonable locality condition. We shall propose a formulation of the stochastic version of Einstein Locality, SEL, which is more satisfactory.[14] Finally, we show that STPI* is entailed by SEL, when the latter is applied to Bell-type situations.

[14] Our formulation is partially based on some remarks by Earman in [41] and in [64].

6.3.1 Bell's formulation of locality in the stochastic case

The history of formulations of locality conditions in the stochastic context seems to begin with Bell in [15]. There he formulates a condition of locality for the stochastic case. He shows that from this condition, one can derive Factorizability*, and hence the BI. If correct, then such an argument supports the claim that Factorizability* qualifies as a necessary condition for locality. Hence, the argument goes, violations of Factorizability* are necessarily violations of locality.

Bell's notion of locality is that given a complete specification of the backward light cone of a given event e, the probability of e should not be modified by information on events located outside the backward light cone of e. According to Bell, this translates in the following way in terms of conditional probabilities:

Definition 28 – ***SL***

Let e be an event located on a region R_e of spacetime. Let σ, stand for the physical state over a time slice S of spacetime across the past light cone of R_e. Let n stand for any event located outside of R_e and its backward light cone. Then SL requires that the probability of the event e, lying over the region of spacetime R_e, be independent of n, conditional on σ:

$$p(e_{R_e}|n, \sigma) = p(e_{R_e}|\sigma).$$

Admittedly, this condition allows for the derivation of the factorization condition. This is because n is interpreted as representing *any* event outside R_e and its past light cone. Indeed, consider two regions of spacetime R_1 and R_2, on which are respectively located the a and b outcome-events. Let λ stand for the physical state over a time slice across the past light cones of R_1 and R_2. Let n, (respectively m), represent any event outside the past light cone of R_1 (respectively R_2). Since n (respectively m) can be any event outside R_1 (respectively R_2) and its backward light cone, it can be, for each side of the experiment, the choice of

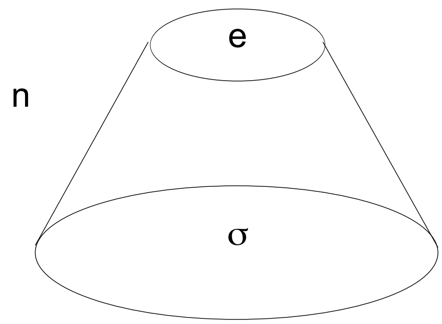

Figure 6.6: Bell's condition of locality in the stochastic case: $p(e|\sigma)$ must be independent of n.

settings or the outcome-event on the other side. Hence, outside of the backward light cone of R_1 are b, j, and eventually δ; while, outside of the backward light cone of R_2 are a, i, and eventually γ. We then obtain, from SL, that[15]:

$$p(a|\lambda, i, j, b) = p(a|\lambda, i), \quad and \quad p(b|\lambda, j, i, a) = p(b|\lambda, j), \tag{6.1}$$

from which in turn one can get the factorization condition. Thus, if SL is the correct construal of locality within the stochastic context, then Factorizability would be derivable from it and hence, qualify as a necessary condition of locality.

Note that, if this were correct, then (PI-LOC)* would be false. Indeed, since Factorizability* is equivalent to the conjunction of Parameter and Outcome Independence* together, failure of Parameter Independence* *or* of Outcome Independence* entail a failure of Factorizability*. If now failure of Factorizability* counted as a case of non-locality, then failure of Parameter Independence* alone *and* failure of Outcome Independence* alone would count

[15]We omit the specification of the regions of spacetime on which the events are localized

as failure of locality as well. In this case, failure of Outcome Independence* would not be less problematic than failure of Parameter Independence in so far as locality is concerned. Thus, the mainstream interpretation is not supported if one follows Bell on the formulation of locality for the stochastic case.

Whether or not the argument above holds depends on whether or not we accept Bell's condition as a condition of locality. Bell takes the spacetime structure as given, and then formulates his locality condition as his "intuitive notion". In the context of this chapter, we can assess Bell's condition within our spacetime framework and our definition of locality as Einstein Locality (Definition 24). The question is thus whether or not Bell's condition qualifies as a satisfactory construal of Einstein Locality in the stochastic context. We shall show in the next subsection that there are good reasons to reject Bell's condition as a satisfactory formulation of Einstein Locality in the stochastic context.

6.3.2 SL is not satisfactory as a stochastic version of Einstein Locality

However natural and intuitive it may look like, SL appears, under closer scrutiny, to be too strong a condition. Suspicion of this comes from the fact that relativistic quantum field theories do not satisfy SL, which Bell himself admits.[16]

The example that Bell gives to the effect that quantum theories, ordinary or relativistic, violate SL is symptomatic of a more serious problem for Bell's condition of locality. Bell mentions an experimental set up in which a radioactive nucleus can emit a single α-particle, and several α-particle detectors are scattered around the nucleus. As long as none of the counters has detected that α-particle, there is a nonzero probability that, say, counter C_i detects the α-particle. That probability becomes zero if one adds the information that, say C_j, has actually detected the α-particle. Bell's locality condition is violated in this case.

[16]See [15, p. 55].

The problem that the example uncovers is that Bell's condition trivializes the notion of non locality because it makes it ubiquitous. Moreover the example shows that the non local correlations identified as such by Bell's condition are certainly not generally those appropriately explainable by a non local physical process between events. So, Bell's condition seems completely inappropriate to the analysis of Bell-type situations.[17] A new locality condition for the stochastic case is required.

6.3.3 Stochastic Einstein Locality

Recall Einstein Locality in the deterministic case: it requires that theories that are Laplacian deterministic be locally so. Laplacian Determinism requires the existence of a Cauchy surface, the state on which determines the outcome-events to the future of the Cauchy surface. Local Laplacian Determinism requires that events on a region R of spacetime be determined by the state on the intersection of a Cauchy surface to the past of R and the backward light cone of R.

Let us propose a formulation of Einstein Locality in the stochastic case by forming natural adaptations of Laplacian Determinism and Local Laplacian Determinism to the stochastic case (yes you read that correctly). The idea is that instead of demanding that the states are determined, we require that probability distributions are determined. We arrive at the following definitions.

Definition 29 – ***Stochastic Laplacianism***

Stochastic Laplacianism requires that any two models which agree on appropriate data on the Cauchy surface agree everywhere to the future of the Cauchy surface on probability distributions of events.

Definition 30 – ***Local Stochastic Laplacianism***

[17] See [64] for a similar criticism.

Local Stochastic Laplacianism demands that the probability distributions over events on a region R to the future of the Cauchy surface be determined by the intersection of the given Cauchy surface and the backward light cone of R.[18]

Definition 31 – ***Stochastic Einstein Locality***

A theory satisfies SEL if when it is Stochastic Laplacian it is locally so.

Note that SEL does not entail SL. This is because it states that, of all possible determining factors for the probabilities on R, only some are in fact relevant. It says nothing about whether there are any correlations between events in R and other events to the future of the Cauchy surface. Again because SEL as formulated above says nothing about correlations, it does not entail Factorizability*, which we make explicit in the following subsection.

6.3.4 SEL entails STPI*, but not STOI*

SEL prescribes that the probability distributions of events localized on a given region R of spacetime be determined by the physical state of an appropriate slice S across the backward light cone of R. How is this applied in the case of Bell-type situations? Consider the same notation as in Section 6.2: a_{R_1} and b_{R_2} for the outcome-events localized on R_1 and R_2, respectively, R_1 and R_2 being two spacelike separated regions of spacetime. Consider then the past domains of influence of R_1 and R_2, and a time slice S across it. Consider then the physical states λ, i, j, γ and δ, are localized on S_1^*, S_2^* and S_3^*.

Concerning a_{R_1}, SEL prescribes that its probability assignment be determined by the physical state on $S_1 \cup S_3$. Since we are assuming that the only possible causal influences are λ, i, j, γ and δ, this region of dependence can be restricted to the subregion $S_1^* \cup S_3^* \setminus S_2$. Similarly, concerning b_{R_2}, SEL prescribes that it be determined by the physical state on

[18] *Note in the above definitions a time asymmetry has been introduced that was not in Laplacian Determinism nor Local Laplacian Determinism, which is natural because we are considering the stochastic case.*

$S_2 \cup S_3$, but this can be restricted to the subregion $S_2^* \cup S_3^* \setminus S_1$. So what SEL prescribes in Bell-type situations for probability assignments on $R_1 \cup R_2$ are determined by $S_1^* \cup S_2^* \cup S_3^*$. It does not imply that correlated events on $R_1 \cup R_2$ should be made independent by conditionalization on the total physical state on S. In other words, SEL does not entail STOI* as defined in 6.2, Definition 25.

If SEL does not prescribe that a be uncorrelated with any event outside the past domain of dependence of R_1 as SL would have it, it still prescribes, once one considers the appropriate time slice S, that a is not determined by the physical state over S_2. And similarly for b. So, in the end, we can say that SEL, applied to Bell-type situations, becomes:

$$p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(a_{R_1}|\lambda_{S_3^* \setminus S_2}, i_{S_1^*}, \gamma_{S_1^*});$$
$$p(b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(b_{R_2}|\lambda_{S_3^* \setminus S_1}, j_{S_2^*}, \delta_{S_2^*}),$$

from which STPI*, as defined in Section 6.2, Definition 26, follows.[19] This should be straightforward when one considers that the equations above are stronger than STPI*: $p(a_{R_1})$ ($p(b_{R_2})$ is made independent of the state on $S_3^* \setminus S_2$ ($S_3^* \setminus S_1$) instead of the state on S_3^* alone. Indeed, STPI* is:

$$p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(a_{R_1}|\lambda_{S_3^*}, i_{S_1^*}, j'_{S_2^*}, \gamma_{S_1^*}, \delta'_{S_2^*});$$
$$p(b_{R_2}|\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}, \delta_{S_2^*}) = p(b_{R_2}|\lambda_{S_3^*}, i'_{S_1^*}, j_{S_2^*}, \gamma'_{S_1^*}, \delta_{S_2^*}).$$

Thus, we have shown that the appropriate version of SEL in our spacetime framework implies STPI* while it does not imply STOI*. As such, STPI* can count as a necessary condition

[19] One might be worried that $\lambda_{S_3^* \setminus S_2}$ may not be well defined since we explicitly left open that it may depend on regions in S_2. In this case, we can simply conditionalize on all λs compatible with the properties fixed in $S_3^* \setminus S_2$.

for locality, while STOI* cannot. In this sense, most of (PI-LOC)* can be supported: failure of STPI* amounts to a case of non-locality while failure of STOI* does not. What cannot be supported is the converse, that is, that STPI*'s holding secures that SEL holds. The defendant of the mainstream interpretation could content himself with this though, if all he wants to sustain is that failure of STPI* are cases of non-locality while failure of STOI* is not. That claim that STPI* is sufficient to secure locality is not needed for the interpretation of the experimental violation of the BI by modus tollens.

Concerning STOI*, we have shown that it is not entailed by our definition of locality within our spacetime framework. Now, at least the three questions arise:

1. If not a locality condition, what is STOI*?
2. Is there no other rationale for taking failure of STOI* as failure of locality in our spacetime framework?
3. Could it be that, while STOI* is not entailed by SEL, the converse be true?

In the following section, we answer 1. in showing that STOI* is equivalent to a spacetime version of the Principle of Common Cause (PCC) when applied to Bell-type situations. This will in turn allow us to answer to 2. and 3. in the negative.

6.4 STOI* is equivalent to STPCC

In this section, we show that the requirement that STOI* holds is equivalent to the demand that the spacetime version of Reichenbach's Principle of Common Cause (STPCC) holds when applied to Bell-type situations. To this aim, we shall present first the PCC (Subsection 6.4.1). We shall then discuss the appropriate spacetime formulation of the PCC (STPCC) and make its relation to STOI* clear. We shall end this section by showing that STPCC is neither a necessary nor sufficient condition for locality, so that STOI* is not either.

6.4.1 The principle of common cause

The PCC is the requirement that given some correlated events, neither of which causes the other, one should be able to recover statistical independence of the events by conditionalization on a common cause. Or, in other words, the dependence between the events should be screened off by a common cause.

Let us illustrate this by way of example. Remember the examples we gave in Chapter 4 in order to illustrate the idea behind the probabilistic theories of causality. More precisely, remember the example of your two cyclothymic friends Jules and Jim.[20] Remember that while Jules and Jim, who do not know each other, display a random distribution of cheerful and crabby days, they are perfectly anti-correlated, one being always crabby when the other is cheerful and vice versa. Remember that there is an explanation where 1. both Jules and Jim have opposite taste concerning the cooking of croissants, 2. they take their breakfast in the same bakery and 3. the baker randomly over- or under-cooks croissants. We can now see that such a situation is a typical example of a common cause pattern: the perfect correlations between two events, none of which causes the other, is explained in terms of a common cause. That is to say, the correlation is screened off when conditionalizing upon the duration of cooking of the croissants in the common bakery. In other word, the Principle of Common Cause is satisfied.

A rigorous statement of the PCC is the following:[21]

Definition 32 – ***The Principle of Common Cause***

If coincidences of two events A and B occur more frequently than would correspond to their independent occurrence, that is, if the events satisfy the relation:

$$P(AB) > P(A).P(B),$$

[20]It should be stressed that Reichenbach's own examples are of the type of our example: macroscopic examples taken from ordinary life situations rather than complicated physical phenomena.

[21]I am here quoting [128, p. 28]. For the original version by Reichenbach, see [43, pp. 158-9].

then there exists a common cause C for these events such that the fork ABC is conjunctive*, that is, satisfies the relations below:*

$$(1)\ P(AB|C) = P(A|C).P(B|C)$$
$$(2)\ P(AB|\overline{C}) = P(A|\overline{C}).P(B|\overline{C}),$$
$$(3)\ P(A|C) > P(A, \overline{C})$$
$$(4)\ P(B|C) > P(B, \overline{C})$$

Conditions (1) and (2) imply that the events A, B and C satisfy the two relations below, called screening off conditions:

$$P(A|B, C) = P(A|C);$$
$$P(A|B, \overline{C}) = P(A|\overline{C});$$

and similarly for A and B interchanged. Reichenbach shows that the conditions (1) to (4) above together imply that $P(AB) > P(A).P(B)$.[22]

The PCC is highly controversial. Many cases of correlations can be sited which do not satisfy it. Further, the four conditions above do not properly characterize common causes. Some examples can be found in which a factor meets all of the above requirements but is not a common cause.[23] Even though the PCC is untenable[24], what the PCC requires is so commonsensical that we seem to use it as a guide (quite successfully) both in everyday life and in scientific research.

[22] [43, p. 160-1].
[23] See [5, note 1] for such an example.
[24] Again, see [5] and references therein for a good overview of the issue.

Reichenbach[25] hoped that the PCC could be part of the characterization of what he calls "normal causality". He characterized normal causality as the requirement that causal effects spread continuously through time.[26] That the PCC be satisfied, that is, that correlations between events which do not cause each other must be explained in terms of a common cause, is, according to Reichenbach, required by "normal causality". By contrast, if there is no such common cause while there are correlated events, then a "causal anomaly" occurs. Normal causality and the requirements associated with it are normative prescriptions for a satisfactory causal picture of the world.[27] Correlations that cannot be explained in terms of a common cause do not satisfy such prescriptions. In other words, the PCC deems pure chance an unsatisfactory explanation for correlations.[28] As said before, the interpretation and status of the PCC are controversial. That all correlations are or should be explainable in terms of a common cause is highly debatable. We shall not further discuss these issues in this dissertation.

For now, it is worth noting that the PCC is formulated without mention of locality. As explained before, probability conditions do not by themselves indicate how the various events and factors are embedded in the spacetime structure. In particular, nothing in the formalism above tells us that the common cause C occurs earlier than its effects A and B. Along the same lines, nothing in the formalism above precludes that we consider the probabilities of C, conditional on A and/or B. That Reichenbach had to add the temporal characteristics of the

[25]We wish to express our gratitude to Alexis Bienvenue for useful discussions about Reichenbach's works and thought.

[26]See, for example, [44, p. 65].

[27]Note that the PCC and normal causality are meant to give metaphysical or only epistemological prescriptions, that is, the issue whether they are meant to describe either the true causal structure of the world or only one of our favorite frameworks when trying to make sense of the world, remains open in Reichenbach's writings. We shall not dig into Reichenbach's work any further in this dissertation.

[28]Pure chance is considered an unsatisfactory explanation in an indeterministic world. Note however that if the probabilities are taken as purely descriptive of frequencies of events in a deterministic world, then it is pre-established harmony, i.e. special initial conditions, that is regarded as an unsatisfactory explanation. See for example [43, p.65].

common causes as restrictions on the above formalism[29] is symptomatic of this problem.[30]

Common causes and their effects thus have to be explicitly embedded within time, independently of the above formalism. In the same way, nothing in the formalism above tells us about the spatial location of A, B and C. Again, the spatial characteristics of common causes and their effects have to be specified independently of the condition in terms of probability distributions. Whether or not the PCC is a locality condition can be assessed only under the condition that such a specification has been made.

Finally, note that, to the best of our knowledge, there is no statement in Reichenbach's own work about the PCC being a locality condition. This historical consideration is in support of the thesis that the PCC, however useful it may be as a methodological principle, is not essentially a locality condition. Before we can produce a rigorous argument to this effect though, we need to discuss the appropriate formulation of a spacetime version of the PCC.

6.4.2 The spacetime version of the PCC and its relation with STOI*

The PCC prescribes that, whenever there are correlations between events which do not cause each other, there exists a common cause, say c, by conditionalization on which the correlations are screened off. How can such a Principle be appropriately applied to our spacetime structure? Let us consider two correlated events e and f. Consistent with Section 6.2, these two events can be located on extended regions of spacetime, say R_e and R_f, respectively. Now, can we choose R_e and R_f such that e and f do not cause each other on the basis of what SEL prescribes? More precisely, does the PCC apply to correlated events

[29] [43, p.162]

[30] Reichenbach hoped to define the temporal order on the basis of the causal order. To this aim, he takes it as an empirical fact that conjunctive forks (common causes patterns) are always open to the future, that is to say, that common causes occur earlier than their effects.

located on spacelike separated regions of a spacetime structure in which SEL holds?

Some clarifications concerning our vocabulary are needed here. It will be recalled from Section 6.3 that SEL prescribes that the probability distribution over events located on a given region R of spacetime be determined by the total physical state depending on an appropriate time slice across the backward light cone of R. More generally, Einstein Locality, whether deterministic or stochastic, imposes some constraints on which regions of space time can *determine* each other. It is common in the literature to refer to such constraints in terms of which regions of spacetime can *cause* each other. On the other hand, the PCC makes use of the term "cause" as well. The so-called "common cause", however, is only characterized by a series of four probability conditions (see Definition 32), without any reference to the notion of physical determination. The term "cause", when used both in the context of the PCC's common cause and of the structure of ordinary relativistic spacetime, is thus ambiguous. Without any further clarification, this could lead to equivocal arguments.

Here is how we shall clarify this situation. First, we shall avoid to talk about common causes altogether. Instead, we shall focus on screening off events. Granted, the full characterization of the common cause includes more than its screening off role. That the common cause satisfies the screening off condition is a consequence of its full characterization. We can focus on this consequence rather than considering a full version of the PCC because it is sufficient for our argument. With this in hand, our formulation of what we still call – with slight abuse of language – the PCC is: there must exists a screening off event for any two correlated events that do not "cause" each other. Now, what "cause" refers to in this formulation of the PCC remains to be filled according to the context in which the PCC is used. In our case, the question arises of how to fill it in the context of a spacetime structure in which SEL holds.

The problem is that it is not all clear whether SEL prescribes anything about causation. This might depend on which theory of causation is considered as well as on which interpretation of relativistic spacetime one holds. In particular, it is not clear that SEL prescribes that events located on spacelike separated events do not cause each other. Since we have not discussed theories of causation yet,[31] and since we do not want to take a specific stance on the interpretation of relativistic spacetimes, let us consider the two options. If one considers that SEL does not impose any constraint on causation, then there is no way to apply the PCC to our spacetime structure. In particular, one cannot argue that the PCC applies to events located on two spacelike separated regions. In this case, the formulation of STOI* is simply unmotivated, let alone motivated by locality conditions.

Let us consider now that the case where the constraints imposed by SEL on which regions of spacetime can determine each other amount to constraints on causation. In this case, SEL is taken to prescribe that events located on spacelike separated regions cannot cause each other. Hence, the PCC applies to any two correlated events e and f located on two spacelike separated regions R_e and R_f. Further, the PCC prescribes that there be a screening off event, by conditionalization on which e and f can be made independent.

Now, what does SEL imply regarding a screening off event? A rather common, but naive, view is that SEL prescribes that the screening off event be located within the intersection of the backward light cones of R_e and R_f. In the case of R_e for example, SEL prescribes that the probability of e be determined by the physical state on an appropriate time slice, say S_e across its backward light cone. Similarly for R_f and S_f. We maintain with Earman that such a view is not supported by relativistic spacetimes.

It will be recalled from Subsection 6.3.4 that, concerning the probability distribution over events on $R_e \cup R_f$, SEL only implies that it is determined by the state on an appropriate slice across the *union* of the backward light cones, $S_e \cup S_f$, and not across the intersection of these, I. In particular, SEL does not imply that the state on a time slice I across the intersection of the backward light cones of R_e and R_f alone determine possible correlations

[31] We shall discuss the issue of causation in Chapter 7.

between events on R_e and R_f. First, consistent with SEL, the future domain of dependence of I is trivial: strictly speaking, it does not determine anything but itself (See Figure 9.5). Second, a synchronization of some events on I is not sufficient to guarantee that events on R_e and R_f be correlated because the correlations could be canceled out by influences which do not register on I, but only on $(S_e \cup S_f) \setminus I$. Third, that we have correlation between events on R_e and R_f does not imply that a synchronization took place between events on I – that is to say, it does not imply that the screening off event lies on I: for I could contain synchronized events itself, and the correlations could propagate without the synchronizing event lying on I. Thus, consistent with SEL, it could be the case that a screening off event does not lie within a time slice across the intersection of the backward light cones of the regions on which the two correlated events are located.

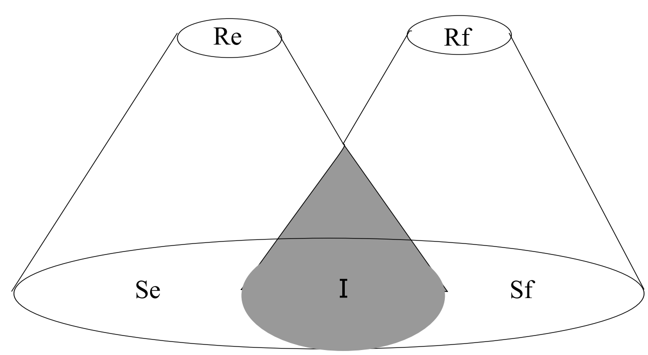

Figure 6.7: The intersection of the backward light cones is trivial.

Note that this is not saying that were some correlations explained in terms of a screening off event lying within their common past, SEL would be violated. Nor is it saying that I cannot be relevant to correlated events on R_e and R_f. In particular, it is not denied above that, if S_e and S_f were fixed, SEL prescribes that variations of the probability distributions on R_e and R_f be determined by variations registering on I. However, the naive spacetime interpretation of the PCC requires more than this determination by I, S_e and S_f being fixed: indeed, it demands that *whatever* S_e and S_f, variations on I *alone* account for synchronization of events on R_e and R_f. This is an empirical question about our world. And it is much more than what SEL can rigorously justify. By contrast, the only statement concerning probability distributions over events on R_e and R_f that SEL by itself implies is that they are to be determined by the total physical state on $S_e \cup S_f$.

In the end then, we can formulate the spacetime version of the PCC:

Definition 33 – ***The Principle of Common Cause: spacetime version – STPCC***

Let e and f be two events located on spacelike regions R_e and R_f of ordinary relativistic spacetime. It is required that, if e and f are not statistically independent, then there is a screening off event c, lying on an appropriate slice across the union of the backward light cones of R_e and R_f, by conditionalization on which e and f are made statistically independent.

$$p(e_{R_e} f_{R_f} | c_{D^-(R_e)\cup D^-(R_f)}) = p(e_{R_e} | c_{D^-(R_e)\cup D^-(R_f)}) p(f_{R_f} | c_{D^-(R_e)\cup D^-(R_f)})$$

Considering the definition above, one realizes that the PCC applied to our spacetime structure (henceforth STPCC) corresponds to STOI*. More precisely, in the case of Bell-type correlations, the requirement above is equivalent to the requirement that STOI* holds. Recall Bell-type situations as described in Section 6.2: a_{R_1} and b_{R_2} are correlated outcome-events located on two regions of spacetime R_1 and R_2. The experiment can be set up so that R_1 and R_2 are spacelike separated. We consider then the backward light cones of R_1 and R_2, and a time slice S across the union of these backward light cones. SEL prescribes that the total physical state σ_S on S determines the probability distributions over events on R_1 and R_2. If there is a screening off event for the correlations between a_{R_1} and b_{R_2}, then it has to be located on S, and conditionalizing on σ_S will guarantee screening off. Conversely, if conditionalizing on σ_S screens off the outcome-events from each other, then there is a screening off event. So, in the case of Bell-type correlations, STPCC becomes the requirement that:

$$p(a_{R_1}, b_{R_2}|\sigma_S) = p(a_{R_1}|\sigma_S).p(b_{R_2}|\sigma_S).; \tag{6.2}$$

In the context we are considering, we have been assuming that the only factors on S that can influence probabilities of a_{R_1} and b_{R_2} are $\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}$, and $\delta_{S_2^*}$. On this assumption, and in this situation then, STOI* as defined in Definition 25, is equivalent to STPCC, which can be seen by substituting $\lambda_{S_3^*}, i_{S_1^*}, j_{S_2^*}, \gamma_{S_1^*}$, and $\delta_{S_2^*}$ for σ_S.

So, we have proved that the requirement that STOI* holds is equivalent to the requirement of STPCC in the case of Bell-type situations. From this, one could object to our conclusion from Section 6.3 that STOI* is not a locality condition in arguing that the PCC is a locality condition, and hence, that STOI* is as well in the case of Bell-type situations. The following subsection is devoted to undermine such objection.

6.4.3 STPCC is not a locality condition

Before we present our objections to it, it will prove useful to analyze the common argument which is given in favor of the claim that the PCC is a locality condition. Roughly speaking, the connection between the PCC and locality is made by saying that the PCC, when applied to correlations between relatively spacelike events, becomes a condition of locality. In Bell-type situations in particular, it seems quite common to believe that, on the one hand, locality is violated because of the failure of common cause models, and, on the other hand, were we to find a common cause, locality would be secured. However natural this idea might seem at first, the argument does not stand when laid out and analyzed in detail.

Let us lay out the argument. Consider two relatively spacelike separated events. Introducing the framework of relativity (albeit without any rigorous formulation), note then that, under a rather common interpretation of relativistic constraints, these two events cannot cause each other: spacelike separation forbids any direct causal link. According to the PCC, if there is any correlation between two given events, this has to be explained in terms of a common cause. But, the argument goes, more can be said about such a common cause. In accordance with relativistic constraints, nothing but events lying within the backward light cone of an event E can have a causal influence on E. If then the common cause is so conceived as to have a causal influence on both of the correlated events, it has to lie within their common past.[32] However, if there is a correlation between relatively spacelike events, but there is no common cause in the common past, then a specific form of causal anomaly would occur, namely: non-locality. That is to say, locality would be preserved if and only if such a common cause could be found. If the above is true, then the PCC, when construed in space-time terms and applied to spacelike separated events is a locality condition.

Anyone would have recognized that the argument above looks very similar to our own analysis in the previous subsection. But in fact, the argument above goes beyond. Let us for now make clear the logic of both arguments, our own in the previous section and the one above, in more detail.

Leaving aside the issues over rigorous definitions and vocabulary, insofar as relativistic spacetimes are assumed to rule out determination of probability distributions other than from appropriate slices across that past light cones, the argument up to (and including) "a specific form of causal anomaly occurs", seems fine. From the requirements of the PCC

[32] It is very likely that anyone formulating this argument would take the "common past" to be within the intersection of the backward light cones. We have already criticized this part of the argument in the previous subsection, and that will not be the point at issue in this subsection. Thus, it is not important what is meant here by "common past".

and of Locality the claim is made that given spacelike correlated events, a common cause must exist in the common past of these events. The problem lies in the very last step of the argument, which states that a lack of such a common cause amounts to a form of non-locality. This last step simply does not follow.

To make this point more clear, let us consider the following notation:

- PCC stands for the requirement that any two correlated events that do not cause each other be screened off from each other by a screening off event;

- SEL stands for the requirement that probability distributions over events located on a region R of spacetime be causally determined by an appropriate time slice across the backward light cone of R;

- $STPCC$ stands for the requirement that any two correlated events located on spacelike separated regions of spacetime be screened off from each other by a screening off event lying within the common past of the correlated events.

Now, the argument above amounts to saying that *if* one takes the PCC to be true, and *if* one wants to apply it to events embedded a space-time structure where Locality holds, then one can conclude that the PCC prescribes that such events be screened off from each other by a screening off condition which lies within their common past. This is perfectly fine and is essentially our own argument in the previous subsection. Given that SEL is the correct notion of locality in our spacetime structure,[33] the argument has the form:

$$PCC + SEL \rightarrow STPCC. \tag{6.3}$$

However, it should be clear enough that, from this, it does not follow that the PCC, nor STPCC is a locality condition. That is to say, it does not follow from this either that (1)

[33]This has been argued in Section 6.3.

violations of STPCC are violations of SEL, or that (2) SEL is secured whenever STPCC is satisfied.

What would be needed to draw Conclusion (1) is that SEL prescribes by itself that correlations between relatively spacelike events be screened off by a screening off event lying in the common past of the correlated events. That is to say, one would have to argue for the following independent premise:

$$^{\dagger}(SEL \rightarrow STPCC)^{34}. \tag{6.4}$$

Only with this in hand could someone argue that a violation of STPCC amounts to a form of non-locality.

What would be needed to draw Conclusion 2, is that SEL cannot fail whenever the STPCC is satisfied, that is to say, one would need to argue in support of the following independent premise:

$$^{\dagger}(STPCC \rightarrow SEL), \tag{6.5}$$

Of course, the trouble is that neither (6.4) nor (6.5) follow from (6.3). Thus, the argument above does not support the thesis that STPCC is a locality condition. What has to be assessed is whether (6.4) and (6.5) hold.

Earman, in [41], shows that satisfaction of the PCC is neither necessary nor sufficient for Einstein Locality to be preserved. We can generalize his arguments to argue that STPCC is neither necessary or sufficient for SEL. First, it is not necessary. The argument here is very similar to our argument that SEL does not entail STOI* in Section 6.3. Indeed, SEL prescribes that the probabilities for the given events be determined by the state on S, but it does not say what form these probabilities should take. In particular, it says nothing about

[34] The dagger indicates that the proposition following is not claimed to be true.

whether the events should be correlated or not. So, STPCC has consequences that are not implied by SEL.

Nor is it the case that STPCC is sufficient for SEL. Earman's counter-example in this case is that the first half of STPCC "can be satisfied in action-at-a-distance particle theories that allow an indefinite number of particles"[35]. We can flesh out this in the following way by appeal to Bohm's theory. That it recovers what is generally taken to be a satisfactory scheme of explanation for Bell-type correlations is a great achievement of Bohm's theory. That it violates Einstein Locality, whether deterministic or stochastic, however, is one of its most well known drawbacks. Thus, Bohm's theory is a non-local theory in which STPCC is secured. Thus, it is simply not the case that STPCC entails SEL.

We shall thus conclude that STPCC neither entails, nor is entailed by SEL. Since STOI* is simply the application of STPCC to Bell-type situations, we can now deduce that the argument to the effect that STOI* is a locality condition on the basis of its relation to the STPCC does not stand. The upshot is thus that STOI* is not a locality condition, but a requirement equivalent to Reichenbach's Principle of Common Cause applied to Bell-type situations. Hence, failure of STOI* is failure of STPCC. The issue of how to interpret failure of the STPCC remains open. In particular, we shall not address the question of whether or not failure of STPCC are to be interpreted in terms of holism in this chapter.[36] What we have obtained though is that (PI-LOC)* is, for the most part, supported, within our spacetime framework.

6.5 Conclusion

We have provided a rigorous spacetime framework, within which we have formulated spacetime versions of Factorizability*, Outcome Independence* and Parameter Indepen-

[35] [41, p. 461]

[36] This will be the burden of Chapter 7

dence* (STFAC, STOI* and STPI*, respectively). We have shown that, within such a spacetime framework, a good case for (PI-LOC)* can be made – or at least for a good part of it. We have defined a rigorous notion of locality by generalization of Einstein Locality to the stochastic case (SEL). We have shown that SEL implies STPI*, when applied to Bell-type situations. Even if the converse does not hold, the defendant of the mainstream interpretation could be content with this result: failure of STPI* corresponds to a case of non-locality.

Further, we have shown that SEL does not entail STOI*, so that failure of STOI* does not imply non-locality. Thus, (PI-LOC)* mostly holds in the spacetime framework proposed here. We closed our analysis by showing that the requirement that STOI* holds is equivalent to the demands of the spacetime version of the PCC (STPCC) when applied Bell-type situations. We have shown that this relation between STOI* and STPCC cannot provide a rationale in order to take STOI* as a locality condition.

At the end of the day, we have thus provided a spacetime framework in which (PI-LOC)* mostly holds. This, of course, does not preclude that other definitions of the spacetime structure considered, or of the versions of the various conditions considered could not yield alternative results. Our result remains that (PI-LOC)* can be supported in a rigorous way.

Chapter 7

The various causal pictures underlying Bell-type phenomena

Recall our earlier definitions:

OI* : For all $a, b, i, j, \gamma, \delta$, and λ,

$$p(a, b|i, j, \gamma, \delta, \lambda) = p(a|i, j, \gamma, \delta, \lambda)p(b|i, j, \gamma, \delta, \lambda)$$

PI* : For all $a, b, i, j, j', \gamma, \gamma', \delta, \delta'$ and λ,

$$p(a|i, j, \gamma, \delta, \lambda) = p(a|i, j', \gamma, \delta', \lambda)$$
$$p(b|i, j, \gamma, \delta, \lambda) = p(b|i', j, \gamma', \delta, \lambda)$$

(PI-LOC)* : Parameter Independence is a locality condition while Outcome Independence is not– locality being understood in a sense linked with Relativity Theory.

(¬PI-CAUS)* : A violation of Parameter Independence* is indicative of an underlying causal relationship while this is not the case for a violation of Outcome Independence*.

(¬OI-HOL)* : (Given that (PI-LOC)* and (¬PI-CAUS)* hold) a violation of Outcome Independence* is indicative of a form of holism.

(¬PI-CAUS)* and (¬OI-HOL)* are statements about whether conditions on probabilistic distributions (PI* and OI*) are causality conditions or not. It will be recalled from Chapter 4, Section 4.3 that analyses à la Jarrett and Shimony can be seen as an investigations of what the hidden causal picture underlying Bell-type phenomena could be. The hope is to infer a causal picture from the probability distributions of the surface phenomena. This presupposes that we are able to construe causation in terms of probability conditions, or at least that the appropriate probability conditions are indicative of causal connections. The main aim of this chapter is to assess whether or not the mainstream interpretation concerning the causal picture underlying Bell-type phenomena can be supported by the theories of probabilistic causation that are available.[1]

In section 7.1, we present the early theories of probabilistic causation, that is, we present theories in which causation was supposed to be characterized by probabilistic notions alone. All theories of probabilistic causation, except for counterfactual theories à la Lewis[2], use the framework of conditional probabilities. We shall thus focus on these theories which use the apparatus of conditional probabilities. We shall explain how the apparatus of conditional probabilities alone is inadequate to characterize causes. Some additional elements have to be added in order to obtain an adequate theory of probabilistic causation.

In section 7.2, we argue that two types of theories of probabilistic causation can be considered. We shall contrast empirical accounts of causation with metaphysical accounts of causation. The former are confined to give an empirical criterion for a relationship between events to qualify as a causal relationship. Empirical theories leave open the issue of the nature

[1]It should be noted from the outset that we do not endorse any of these theories of causation. We use them as frameworks for a rigorous assessment of the mainstream interpretation.

[2]Two loci classici for Lewis' account are [80] and [81]. As mentioned in Chapter 4, we shall not deal with counterfactual theories of causation.

of the objective counterpart at the ontological level of what we take as causal relationships at the empirical level. By contrast, metaphysical accounts of causation are designed to give an account of causation as an objective feature of the world. Such accounts typically take causal processes and causal interactions, which are embedded in spacetime, as fundamental.

The most natural type of theory of probabilistic causation seems to be manipulability theories of probabilistic causation (henceforth MTPC), which are of the empirical type. That said, we shall argue that spacetime theories of probabilistic causation (henceforth STPC) are also good candidates to give an account of probabilistic causation, namely a metaphysical one.

In Section 7.3, we turn to assess whether or not (¬PI-CAUS)* and/or (¬OI-HOL)* hold first within MTPC, and second within STPC. We shall argue that within MTPC, a rigorous argument can be made in favor of (¬PI-CAUS)*. That said, because MTPC are only concerned with causation from an empirical point of view, MTPC can support (¬PI-CAUS)* only in a weakened sense: whereas failure of PI* are indicative of causal relation which have empirical manifestations, failure of OI* is not indicative of any causal relation which has empirical consequences. Further, because of this empirical stance, MTPC does not support any specific interpretation of failure of OI*.

Within STPC, we shall argue that in general the mainstream interpretation fails. Under a strict application of the theory, either failure of PI* and failure of OI* are both indicative of a causal relation, or neither are. That said, we shall give some plausibility arguments in favor of a weakened version of the mainstream interpretation, in the sense that it has strong epistemological, and not metaphysical leanings. The lesson of Bell-type experiments, given this weakened version of the mainstream interpretation, is more about how we should conceptualize the world than what the world is like at the fundamental level.

The upshot of the chapter is:

1. None of the existent theories of probabilistic causation support the mainstream inter-

pretation understood as strong program of experimental metaphysics;

2. MTPC provides a rigorous framework to support an empirically oriented version of (¬PI-CAUS)

3. Refined STPC provide plausibility arguments for the interpretation of failure of OI* as in (¬OI-HOL)*.

7.1 How early theories of probabilistic causation fail

In this section, we present early theories of probabilistic causation and what they were supposed to do. We then turn to explain how they fail to characterize causes adequately.

7.1.1 Early theories of probabilistic causation

Reichenbach[3] is one of the fathers of theories of probabilistic causation. Good and Suppes[4] have both provided more elaborated accounts of Reichenbach's ideas. All three intended to *reduce* the notion of causation to probabilistic notions. That is to say, they took probabilistic relationships between events to be fundamental: causation was to be characterized by probabilistic notions alone.[5]

A reductive stance is not required to admit that probabilistic notions are necessary for an account of causation under the hypothesis that the world is irreducibly indeterministic. While the reductive program has been abandoned by most philosophers,[6] the discussion over theories of probabilistic causation have developed at a high speed in the literature.[7] The

[3]See in particular [95].

[4]See [58, 59] and [114].

[5]Such reductive theories have been convincingly criticized. See, among others, Salmon, for example in his [96], reprinted with some corrections in [99, chap. 14].

[6]Hitchcock is a particularly good example [136].

[7]The most robust contributions are by Spirtes, Glymour and Scheines in [107], and Pearl in [88]. The most recent developments in the field are due to Hitchcock and Woodward ([137, 138]).

majority of these authors take the further step toward MTPC. That said, Salmon, and Dowe after him, have developed another view, of the spacetime type, on the matter.[8] Salmon for instance considers the use of the probability apparatus to be necessary, albeit not sufficient, to give an adequate theory of causation.[9]

In Subsection 4.3.1, we have given the probabilistic characterization of cause and causal relevance:[10]

Probabilistic cause: e is a cause of f if and only if

$$p(f|e) \geq p(f|\neg e) \tag{7.1}$$

If the probabilities are non zero, then this is equivalent to:

$$p(e, f) \geq p(e).p(f) \tag{7.2}$$

Probabilistic causal relevance: e is a causally relevant to f if and only if

$$p(f|e) \neq p(f|\neg e) \tag{7.3}$$

If, again, the probabilities are non zero, then this is equivalent to:

$$p(e, f) \neq p(e).p(f) \tag{7.4}$$

There are three notes of clarification concerning the use of the probability calculus for accounting for causation which are useful for our purposes.

First, probabilities are defined for events. Typically various things are said to be causes, not only events, but also entities, properties, or changes. Since we are using the formalism of conditional probabilities (by contrast to conditional counterfactuals for example, which avoid such a problem), and since the condition above is perfectly symmetric (as is clear from (7.2) and (7.4)), not only the effects but also causal factors must be capable of being assigned

[8]See [99, chap. 13] and [38, 39].

[9]in [99, chap. 14] for example.

[10]See Definitions 12 and 13, the latter accommodating the possibility of negative relevance.

probabilities in our framework. Whatever the kind of causes we consider, it is thus required that they can be construed as events which either occur or do not occur.

Second, the probability of f given e is not defined if the probability of e is zero. One way out of this is to take conditional probabilities as fundamental, and to define the single probabilities in terms of conditional probabilities.

Finally, notice that events, as they appear in probability conditions, are not located in spacetime. Thus, if one wants to consider causal factors and events as embedded in spacetime, one has to add additional structure to the above characterization. In particular, since, again, the conditions above are perfectly symmetric, there is no indication of which event among e and f occurs before the other.

With this in hand, we can now explain why the characterizations of causation in terms of probabilistic notions alone fail in general.

7.1.2 The failure of accounts of causation in terms of probabilistic notions alone

There are classical criticisms leveled against accounts in which causation is characterized in terms of probabilistic notions alone. In this subsection, we investigate some of these counterexamples in order to draw a list of desiderata for any adequate theory of probabilistic causation. In short, the desiderata will be: 1. a clear specification of the probability space; 2. some additional structure which introduces an asymmetry between cause and effect.[11]

In the literature, the main criticism leveled against the above characterization of causes (Equations (7.1) to (7.3)) is that they are neither necessary nor sufficient to characterize causes. That it is not *sufficient* to raise the probability of an event to count as a cause is exemplified by the classical type of example utilizing spurious causes,[12] which appear in

[11]This last stipulation should not be made by fiat, to rule out something like backwards causation or the like.

[12]I borrow here and throughout the chapter Suppes' vocabulary in his [114]. He distinguishes between

common cause situations.

Consider the following counterexample. Arguably, it is statistically the case that people who have yellow fingers tend to suffer from lung cancer more often than average. Following the pattern above, this would seem to indicate that to have yellow fingers is a cause of lung cancer. Of course, we all realize that this is a spurious cause: in fact, yellow fingers, as well as lung cancer are taken to originate in a common cause: heavy smoking. This seems pretty obvious and easy enough to figure out. More tricky situations could occur though. For example, we believe that smoking is a cause of lung cancer. Now, it might be the case that a certain gene provokes lung cancer in the presence of smoke, while smoking by itself, without the gene, would not enhance the chances of getting lung cancer.

The criteria given by Equations (7.1) to (7.3) do not give us a way to reject spurious causes of the foregoing type. Arguably though, any satisfactory theory of causation should give us the means to distinguish between spurious and genuine causes. Some further conditions are needed. The issue of which kinds of amendment or complementation of the characterization above could provide a satisfactory answer to the challenge of spurious causes will be addressed later on. For now, let us show by way of counterexamples that the characterization of cause above is not necessary either.

Typical counterexamples are related to Simpson's paradox. Simpson's paradox occurs when there are some partitions of a given population are such that while an event e_1 has a higher probability than e_2 within every subset of the population, it is not the case at the level of the total population. A classical example is that the partition of the hiring offerings and of the gender of the candidates actually hired in a given university can be such that, while women are more likely to be hired at the level of each department, the opposite is true when one looks at the statistics at the level of the university, that is, of all the departments

---

1. prima facie causes, 2. spurious causes, and 3. genuine causes. Prima facie causes are candidate to qualify as a cause. Spurious causes are prima facie causes that are found out not to be genuine causes.

taken together.

Applied to causal patterns, instances of Simpson's paradox can lead to cancellations of correlations. Here, a classical example is the example of running, smoking and heart attack. Arguably, smoking is a cause of heart attack. On the other hand, running on a regular basis is a cause of heart strength, which protects from heart attacks. Now, it might be the case that, in a given population, smokers are also more likely to be regular runners (smoking and running being both means to fight anxiety for example). The partition could be such that, in the total population, smokers are less likely to suffer from heart attacks than the average non smoker. Hence, it is not necessary that conditioning on a cause raises, or even changes (in case of a perfect cancellation of correlation) the probability of its effect.

In both types of counterexample above, an important problem resides in that some other events interfere within the causal relationship considered. The choice of which factors are considered, in other words, the choice of a probability space, is of great importance for an adequate assessment of probabilistic causation. If one takes into account the smoking factor in the first example, and assess the relationships between the yellow fingers and lung cancer, conditional upon heavy smoking, the problem dissolves. Compare the smoker-runner to the non-smoker runner instead of the average non smoker, and the theory recovers the causal relationships as expected. A first requirement for any adequate theory of probabilistic causation is the means to specify adequately which factors have to be taken into account.

Once the probability space has been adequately specified, then additional structure can be added to Equations (7.1) to (7.3), so that we can sort out genuine causes from spurious causes. The intuition is that we need to isolate the causal relationship being considered from interfering factors. This intuition is codified by the following:

(ISOLATION) : an adequate theory of probabilistic causation should include means to discard the influence of factors other than the putative cause on the putative effect under consideration.

One structure that satisfies (ISOLATION) is the famous no screening off condition. The no screening off condition is part of the Principle of Common Cause formulated by Reichenbach.[13] The no screening off condition states that, given that an event e is a candidate for being a cause of an event f (the probability of occurrence of f, given that e occurred, is higher or at least different than if e had not occurred), e is a genuine cause of f only if there is no other event c, conditional upon which the correlation between e and b is canceled.

Definition 34 – ***Genuine Cause***

An event e is a genuine cause of an event f if and only if:

1. e is causally relevant to f, and

2. there exist no event c such that:

$$p(e, f|c) = p(e|c)p(f|c).$$

On the other hand, if there is such an event c, then c is said to screen off e from f, and e is considered to be a spurious cause:

Definition 35 – ***Spurious Cause***

An event e is a spurious cause of an event f if and only if:

1. e is causally relevant to f, and

2. there exists an event c such that:

$$p(e, f|c) = p(e|c)p(f|c),$$

in which case, c is said to screen off e. 2. is called the no screening off condition.

We may refer to such a c in the above definition as the screening off event in the following.

Now the question arises: does the addition of the no screening off condition provide satisfactory to incorporate the foregoing types of counterexamples? The no screening off

[13] in [43, 44]. See 6.4.1 for a more detailed account of the Principle of Common Cause.

condition is often taken to solve the first half of the challenge above, that is, the issue of sufficiency of the probabilistic criterion for qualifying as a cause. Remember the example of yellow fingers and lung cancer. It seems clear that in this case, the correlations between having yellow fingers and suffering from lung cancer is screened off by an appropriate event, indeed, heavy smoking. Hence, the no screening off condition seems to allow us to reject some causes as spurious.

The no screening off condition might be sufficient to get rid of the cases of spurious causes of the common cause pattern type; however, the problem necessity regarding the conditions given in Equations (7.1) to (7.3) remains. In other words, the screening off condition does not solve the problem of cancellation of correlations. However, the intuition behind the no screening off condition, that is, to study the relationship considered, while holding still any interfering factors, seems to remain the right one. The main problem in our example of cancellation of correlation is that it so happens that there is "more that one way to skin a cat". Hence, if we want to investigate the causal structure of the world, we have to isolate the various causal relationships (or the various causal processes) from one another. What we need is a more restrictive and more complete way to specify the situation in which the relationship can be assessed.

However, we would like to emphasize that theory of probabilistic causation will be adequate only if some additional structure that introduces asymmetry is added. This is because while a condition that elaborates on the no screening off condition in order to satisfy (ISOLATION) above generally will allow one to identify genuine causal relations between events, it still does not tell us which of the events is the cause and which is the effect. Taking the example of the no screening off condition, if we cannot find any event that screens off lung cancer from heavy smoking, the probabilistic apparatus does not have enough structure to tell us which causes which. Any appeal to a requirement of the type "causes happen before their effects" would preserve the reductive analysis, for it precisely amounts to introduce some additional structure, namely the asymmetry of the time order. Hence, besides (ISOLATION), a second desideratum for any adequate theory of probabilistic causation is that additional structure be added which introduces some asymmetry in the model:

(ASYMMETRY) : an adequate theory of probabilistic causation should include a structure introduces some asymmetry in the model.

7.2 Two types of theories of probabilistic causation

In this section, we argue that both manipulability theories and spacetime theories of probabilistic causation (MTPC and STPC) meet the minimal desiderata distinguished in the previous section. Both MTPC and STPC are thus good candidates for qualifying as adequate theories of probabilistic causation.[14] We shall first explain how MTPC and STPC fall under two different categories of theory of causation (Subsection 7.2.1). We shall then explain why MTPC and STPC are quite natural candidates in Subsections 7.2.2 and 7.2.3 respectively.

7.2.1 Two types of theories of causation

To put it roughly, MTPC and STPC correspond to an empirical and a metaphysical stance, respectively, over what counts as a satisfactory theory of causation. Theories of the empirical type, henceforth type (M), are designed to give an empirical criterion for an event A to qualify as a cause of an event B. Causation in this case is construed as a relation between events. The world is conceived as a set of events located in spacetime. We navigate such a manifold by arranging events in types and in regularities in relations between events,

[14]Note that we are not saying that such theories are without flaws or difficulties. It is not our point to take a stance over which theory of probabilistic causation is the best either. Rather, we want to point out that MTPC and STPC fulfill the minimal desiderata for any theory of probabilistic causation to be viable. To fulfill such desiderata is necessary, but we do not claim that it is sufficient.

or between types of events. Certain events, or types of events, are always or most of the time followed by certain other events, or types of events. Some of these regularities we construe as causal relationships. For instance, if every time you put your hand in fire, you find your hand (probably painfully) burned, then you will construe the event types 'hand in fire' and 'hand burned' as linked by a causal relationship of the type: (for human hands at least), 'fire causes burns'.

Of course, not all regularities qualify as causal relationships. For example, that almost every day my arrival on the track of my morning train is followed by my neighbor's arrival on the same track is not indicative of a causal relationship between these two events. Rather, it is just the case that my neighbor and I need to catch the same train to be at work on time. Moreover, I tend to be more anxious to miss the train than he is, so that I take always more extra time than he does. Even more simply, that every morning my washing my face is followed by my brushing my teeth is not indicative of a causal relationship either. Based on the sole observation of regularities, fire, my arrival on track B, and my washing my face, are candidates for qualifying as causes. However, while fire seems to count as a genuine cause, both my arrival on track B and my washing my face seems to be spurious causes. Theories of causation of type (E) are designed to give adequate and workable criteria for distinguishing between genuine and spurious causes.

Within such theories, the question whether causal relationships are ontologically implemented is at least left open, if not answered negatively. It is well known that the suspicion over the construal of causal relationships as grounded on real physical connections between events is coming from Hume's analysis of causation. Hume indeed famously argued that no effect can be logically deduced from its cause. He suggested we have no satisfactory reason to believe that what we call causal relationships correspond to any physical connections at the ontological level. Theories of type (E) are in line with Hume's skepticism. It leaves room for a form of agnosticism as to the ontological status of physical connections corresponding to what we call causal relationships.

The second half of the twentieth century has seen the resurgence of the second type of theory of causation, henceforth theories of type (M), in which causality is accounted for at the ontological level. Salmon[15] has initiated such a revival, in claiming to have an account of these physical connections that Hume could not find. Salmon thus proposes an analysis of causality from a metaphysical stance. Causality is accounted for not in terms of relations between events, but rather in terms of causal processes and causal interactions, which Salmon takes to be both objective and contingent. Salmon and his followers insist on the fact that an account of causality is not satisfactory if it is confined to giving an empirical criterion for the detection of causal relations. Instead, a satisfactory theory of causality must give an account of what causality is at the ontological level.

One rationale for taking theories of type (E) to be unsatisfactory is the following. It is arguably an important feature of causal relationships that they provide warrant for our inferences. In other words, to construe the world as a manifold of events interrelated by causal relationships helps us navigate such a world. Typically, after a few instances of burning your hands, you will probably stop putting them in fire (however tempting it might be). Every day life as well as scientific activity feature numerous cases of such inferences. It should be clear that the theories of type (M), in construing causation as corresponding to physical connections, provide better warrant for our inferences than theories of type (E).

However, not all accounts of type (E) would maintain that empirically distinguishable regularity relations are all there is to causality. In other words, not all defendants of theories of causation of type (E) would attempt to reduce causality to empirically distinguishable relationships between events. Some theories of this type instead rely on a basic notion of probabilistic cause, or of causal mechanism. These, however, are taken as fundamental. Thus a further, metaphysical account of such a fundamental notion is not possible.

[15] [97], [99].

It should be clear that the original theories of probabilistic causation, as formulated by Suppes and Good, are of the empirical type. So, it could seem natural to build on the original insight and develop better theories of probabilistic causation of the empirical type. MTPC are indeed of such type. By contrast, Spacetime theories of causation are of the metaphysical type. So, manipulability theories appear as the most natural way to extend earlier account of probabilistic causation to make them adequate.

7.2.2 Manipulability theories of probabilistic causation

There are basically two kinds of manipulability theories of causation. One kind contains theories in which causation is accounted for in terms of the free action of an agent. The second kind contains theories in which causation is accounted for in terms of (non-human, and neither necessarily actual nor necessarily physical feasible) intervention. The former has been advocated by Von Wright[16] and, more recently, Menzies and Price.[17] The latter has been advocated by Pearl[18] and, more recently, by Hitchcock and Woodward.[19] The main criticism that is usually leveled against the former kind of theory is a form of "anthropomorphism". We shall not enter the debate, even less pretend to settle it. In the context of this chapter, we shall consider Hitchcock and Woodward's theories in which causation is accounted for in terms of interventions, to which the objection of anthropomorphism does not apply. We shall argue that interventionist MTPC meet the two desiderata we distinguished in Subsection 7.1.2, namely (ASYMMETRY) and (ISOLATION).

According to interventionist MTPC, a causal relationship holds between *variables*, which can take a spectrum of possible values. We shall keep our previous notation, in which capital letters are used for variables, X, Y, and small letters for the various values $\{x, x'\}$, $\{y, y'\}$

[16] [133]
[17] See [84], [89].
[18] See [88].
[19] in [137, 138], [136].

these variables can take. Notice that a variable is only said to be a cause *with respect to* a set of variables. This can be taken as an expression of epistemic modesty. As an empiricist, one can never be assured that one has ruled out all possible causal influences, hence the interactionist view hedges their bets by making judgments of causal relationships relative to the variables considered. Once the set of variables under consideration is fixed, a necessary and sufficient condition for some X to qualify as a cause is:[20]

Definition 36 – ***Cause: manipulability account***

A variable X is a cause of a variable Y, with respect to a set of variables $\mathcal{S}$, if and only if the value of Y undertakes invariant changes with some intervention on the values of X.

Let us explain the notions contained in this definition. First, the notion of invariant change of the values of the variable Y corresponds to the following idea: there are some interventions on the variable X such that the amount of change obtained in Y is related to the amount of change imposed on X. More precisely, a certain relation remains invariant under the changes of the values of X. A typical example would be the Coulomb Law. Consider a circuit in which you have put a generator and a resistor. You want to check whether the generator is a cause of the current in the circuit. According to the interventionist manipulability theories of causation, the generator counts as a cause of the current if a change in the voltage associated with the generator results in a change in current, *and* that such changes follow an invariant law, in this case the Coulomb law $\Delta V = R\Delta I$.

Second, the crucial notion of intervention is defined in the following way:

Definition 37 – ***Intervention***

An intervention I on a variable X with respect to a variable Y consists in changes in the value of an intervention variable for X with respect to Y, where the notion of an intervention

[20] See for example the concise presentation of these concepts in [137, p. 12-13]. For a more complete account, see [136, chap. 2-3]. The formulations of the definitions we give are paraphrasing one or the other.

variable I is characterized by four conditions:

1. *I causes X;*

2. *I can switch off other causes of X;*

3. *I does not have any effect (either as a direct or as an intermediary) on Y, outside the causal path which goes from I to Y through X;*

4. *I is statistically independent of any variable Z that causes Y through a causal path which does not go through X.*

One will notice that the notion of intervention is clearly a causal notion and the account of causation given by MTPC is circular. But the circularity is not vicious as they do not aim to give a reductive account of causation. Instead, they aim to give an account of how one can identify whether a relationship is causal or not. A basic notion of cause and some knowledge of some of the causal relationships among the variables considered is required for assessing whether two particular variables, A and B, are causally related. That said, none of such pre-required knowledge concerns A and B themselves nor their relationship. The account would be viciously circular if, in order to assess whether or not A is a cause of B, one would have to know whether or not A is a cause of B. This is not what is needed here. All pre-required knowledge concerns only other causal relationships than the one under study. So we set this problem aside.

The role of the four conditions above can be explained in the following way. Conditions 1. and 2. are constraints which guarantee that X is affected by and only affected by I. Conditions 2. and 3. are constraints such that I affects Y only through X. That is to say, there is no route from I to Y other than the route that goes through X. Let us clarify this with an example.

The favorite example of defendants of interventionist MTPC is the case of a medical study. Typically, the volunteers for the study are divided into two groups: the members of one group are given some medication, while the members of the other are offered a placebo. In such a study, giving or not giving patients medicine corresponds to the intervention I. Patients processing the medicine is treated as the putative cause, X. Patients recovery is the effect under investigation, Y. It is easy to see that the constraints imposed by conditions 1. to 4. are satisfied. We can illustrate the first type of constraint (corresponding to conditions 1. and 2.)– that the putative cause be affected and only affected by the intervention – is satisfied as follows. The study is set up so volunteers do not have independent access to the medicine during the study. Also, volunteers are sought that lack any special physical condition so that the medicine given to them cannot be processed by their bodies. In short, we want our intervention to be the only cause of the medicine being processed or not processed by their bodies.

The second type of constraint – that there be no other route from I to Y than through X – can be illustrated as follows. The volunteers' recovery (or non recovery) should depend only on whether one has processed the chemical substances contained in the medication. Notice that we do not do medical studies by simply dividing between people who take the pill, and people who do not take anything. This is because we know that the very fact of taking a pill (whether it contains active medication or not) does influence the chances of recovery. Having the people in the non-medicated group take a placebo neutralizes such an interfering factor. Moreover, volunteers for the study are randomly distributed into the test group and control group which ensures that other interfering factors, such as resistance to the illness in question, are statistically independent of the intervention. These techniques ensure that an intervention affect the effect only through the putative cause under study (as understood in 3. and 4. above).

It should be clear that the apparatus of the intervention as described above makes interventionist MTPC fulfill (ISOLATION). The apparatus of the intervention indeed allows one to define a context in which a well controlled assessment of the causal relationship under consideration can be accomplished.

Let us now see whether or not MTPC can satisfy (ASYMMETRY), that is, provide additional structure that introduces asymmetry required in any adequate theory of probabilistic causation. The apparatus of interventions actually does introduce the required asymmetry. This is because the very notion of intervention is asymmetric, as is the criterion in terms of intervention: one can change an effect by manipulating the cause, while one cannot change a cause by manipulating an effect. So, in the end, interventionist MTPC do fulfill the two desiderata for an adequate theory of probabilistic causation.

Let us see how this works on our examples. Consider the problem of spurious causes first, which was illustrated by the example of heavy smoking, yellow fingers and lung cancer. MTPC render this objection inert by extending sets of variables considered rather than making bald judgments about whether a variable is a cause or not; one checks whether a variable is a cause or not. Second, the method of determining causes would require that the factor "heavy smoking" be held still, so that our intervention on the color of the finger be the only potential influence on lung cancer. One would quickly discover that the color of the finger is not a cause of lung cancer relative to the set of variables that considers heaving smoking as well. Finally, the asymmetry introduced by the notion of intervention allows us to determine which of the variables are causes and which are effects, with respect to our set of variables. Indeed, dying a smoker's fingers, making him or her use a cigarette holder will not (neither significantly nor invariantly) have any effect on whether or not he or she contracts cancer. Similarly, implementing lung cancer in a volunteer will not in general make him smoke heavily. By contrast, making him or her stop smoking will have a preventative effect on both his or her getting lung cancer and yellow fingers (both significantly and invariantly). So, arguably, MTPC seems to be a great improvement over earlier theories of probabilistic causation. We may plausible consider it a potentially adequate theory of causation.

Let us turn to the issue of the cancellations of correlation now. The problem was that the partition and the probability distributions on a population can be such that the average smokers are less likely to suffer from heart attack than the average non-smokers for smokers that are also more likely to be regular runners. Hence on early theories of probabilistic causation we would not recognize the true causal factors at play. It seems that in considering the running factor in addition to the smoking factor for heart attacks – we would have to check first that running is a cause of heart strength – and holding fixed the running variable when studying the smoking variable we would easily arrive at the correct conclusion via interventions. So, if the condition of interventions is fulfilled, it seems that MTPC allows one to discover the correct causal structure in the set of variables considered.

In short then: interventionist MTPC arguably qualifies as a minimally adequate theory of probabilistic causation. Let us turn to STPC.

7.2.3 Spacetime theories of probabilistic causation

In this subsection, we argue that spacetime theories of probabilistic causation (STPC) are good candidate for an adequate theory of probabilistic causation. That is to say, we argue that such theories fulfill the two minimal desiderata distinguished in Subsection 7.1.2: (ASYMMETRY) and (ISOLATION).

Let us recall first the main tenets of STPC. The main theories here are Salmon's and Dowe's. Salmon[21], and Dowe[22] give an account of (meta)physical causation in terms of causal processes and causal interactions.[23] Now, it will be recalled that the Humean tradition takes it that we should be suspicious about any appeal to causal powers, or physical causal connections. Regarding such tradition, Dowe is more cautious than Salmon for he only aims at giving only an empirical theory of physical causation. By contrast, Salmon claims that

[21] To begin with, see [97, 99, 98]. Note that ([98]) is a response to criticisms leveled by Dowe and Kitcher against ([99]).

[22] See [38]

[23] Causal interactions are defined in terms of causal processes, so that causal processes are really the most fundamental notion in STPC.

causal processes, as characterized in his theory, are precisely the causal connections that Hume was unable to find.

In the case of probabilistic causation, both authors reject the characterization in terms of probabilistic notions alone. Causal relationships between events, whether or not discoverable by surface correlations, are not what constitute the causal structure of the world. More fundamental are causal processes. When investigating the causal picture underlying some phenomena, the main challenge for STPC is to give an adequate characterization causal processes, that is, a characterization that allows to distinguish between genuine causal processes and "pseudo processes". An example of this distinction would be that a moving car is a causal process, whereas its shadow on the ground is a pseudo process.

Salmon and Dowe disagree on the adequate criterion to distinguish causal processes from pseudo processes. Salmon, in his original theory,[24] took it that causal processes are characterized by their capacity to transmit marks.[25] Dowe[26] leveled important criticisms against Salmon's mark method, and developed his own account. Dowe claims that what characterizes a causal process is the possession a conserved quantity. Conserved quantities are quantities governed by conservation laws. Typically, particles, waves, or dogs in movement possess a conserved quantity, whereas shadows do not. Salmon[27] has adopted Dowe's account in terms of conserved quantity (thus abandoning the mark method), but remained firm on the idea that a notion of transmission is indispensable for an adequate theory of causal process.

According to Salmon, to possess a conserved quantity is not sufficient to qualify as a genuine causal process: the ability to transmit (by itself) a conserved quantity is essential

[24]in [97].

[25]The "mark method" was first formulated by Reichenbach. Salmon borrowed the concept but provides a much more sophisticated account.

[26]See [36, 37]. Kitcher and Hitchcock took also part to a fascinating discussion that lasted at least a decade.

[27][99, 98].

to causal processes. Consider Salmon's classical example of the pseudo process constituted by a spotlight moving on a wall. Now consider the entity, called it the "blob", made by the various parts of the wall which are successively enlightened by the spotlight. It should be clear that the blob possesses a conserved quantity, namely energy. But surely an adequate theory of causation should identify the blob to a pseudo process. Salmon rejects the blob as a causal processes because it does not transmit energy over its history. Instead, it is the source of the light that transmits energy to the blob at every moment of time. Hence, the beam of light from the source to the wall counts as a causal process, while the blob does not.

Dowe, on the other hand, rejects the necessity of a transmission criterion. Roughly speaking, the role played by the notion of transmission in Salmon's theory is replaced by a primitive notion of an object in Dowe's theory. According to Dowe, only objects can be causal processes. Causal processes are objects that possess a conserved quantity. Thus, the blob does not count as a causal process for it does not count as an object in the first place. How do we recognize an object? Objects are to be distinguished from "spatiotemporal junk".[28] An object is defined as being wholly present at any given moment of time of its history. As Salmon notes in [98], this amounts to define an object as genidentical. In the example above, the blob does not count as an object for the various part of the wall are not made of the same molecules.

Both Salmon and Dowe have advanced good arguments and clever examples in order to defend their respective theories. We shall not enter the debate, even less pretend to settle it. That said, in the context of this chapter, we shall use Salmon's last version of his theory instead of Dowe's. Our reason for this is the following. It is not clear that the definition of objects in terms of genidentity is tenable in general. Most of all, it is highly doubtful that it can be used consistently as a primitive notion for quantum phenomena. Thus, we shall use Salmon's account of causal processes and causal interaction in terms of, respectively,

[28]The phrase is Kitcher's.

transmission and exchange of conserved quantity.

Salmon defines causal processes and causal interactions as follows:[29]

Definition 38 – ***Causal Interaction***

A causal interaction is an intersection of world-lines that involves exchange of a conserved quantity.

Definition 39 – ***Causal Process***

A causal process is the world-line of an object that transmits a non-zero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).

Definition 40 – ***Transmission***

A process transmits a conserved quantity between A and B ($A \neq B$) if and only if it possesses [a fixed amount of] this quantity at A and at B and at every stage of the process between A and B without any interactions in the open interval (A, B) that involve an exchange of that particular conserved quantity.

In the case of probabilistic causation, the interesting function of causal processes is to transmit determinate probability distributions.

> Transmission of a determinate probability distribution is, I believe, the essential function of causal processes with respect to the theory of probabilistic causality. Each transition event can be considered as an intersection of causal processes, and it is the set of probabilities of various outcomes of such interactions that constitute the transmitted distribution.[30]

An example will serve to clarify what "transmission of a determinate probability distribution" is.[31] Consider an atom of a radioactive element has some probability to decay. Each

[29]The definitions given are quoted from [98, p. 462-468].
[30][99, 228].
[31]We take our example from Salmon himself.

transition to another level counts as an interaction: the photon being either emitted or absorbed is the other causal process with which the atom, as a causal process, interacts. In between two transitions, the atom "carries" with it the probability distribution over possible transitions.

Causal processes and causal interactions have different roles within the account of the causal structure of the world given by STPC. On the one hand, causal processes are responsible for the *propagation* of causal influences. It is essential in STPC to consider that causal influences travel continuously in spacetime. Causal processes are thus the fundamental building blocks of the causal structure of the world. On the other hand, causal interactions are responsible for the *production* of the causal structure of the world. That is to say, causal interactions are responsible of the modification of the causal influences that causal processes propagate. There are two fundamental kinds of causal interaction, the structure of which is given by interactive forks and conjunctive forks, respectively.[32]

Conjunctive forks are defined as follows:[33]

Definition 41 *A conjunctive fork between events A, B and C is characterized by the following four conditions:*

1. $P(A.B|C) = P(A|C).P(B|C)$,

2. $P(A.B|\neg C) = P(A|\neg C).P(B|\neg C)$,

3. $P(A|C) \geq P(A|\neg C)$,

4. $P(B|C) \geq P(B|\neg C)$.

[32] We do not mention perfect forks, because they are limit cases of either conjunctive or interactive forks. See [99, p. 297 sq.] for more details.

[33] Given our framework, in which causal relevance in general is given the conceptual priority over positive causal relevance, we shall feel free to replace the two last conditions by simple inequalities.

Such a definition, of course, corresponds to Reichenbach's own definition of conjunctive forks.[34] Typical illustrations of conjunctive forks are common cause patterns, such as our example of your two cyclothymic friends Jules and Jim and their rather scattered common baker in Subsection 4.3.1.

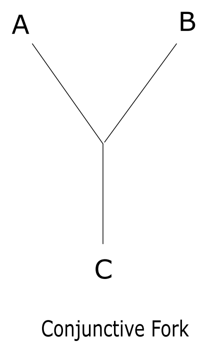

Figure 7.1: Conjunctive forks produce the causal structure of the world.

Interactive forks consist in the intersection of two causal processes, with an exchange of a conserved quantity. Typical examples of interactive forks is Compton scattering[35], or collisions of billiard balls.[36] The structure of interactive forks is very similar to that of conjunctive forks, except that the common cause does not screen off the correlated events. That is to say, the correlated events A and B are not made statistically independent by conditionalization on the common cause C. Instead, we have an inequality instead of an equality:

$$P(A.B|C) > P(A|C).P(B|C). \tag{7.5}$$

[34] [43].
[35] See [99, p.133].
[36] See [99, 293].

Note that the structure of the conjunctive forks is more appropriate to Bell-type situations. In the following, we shall not make much use of interactive forks. Figure 7.1 illustrates conjunctive forks.

In short then, the causal structure of the world is constituted by causal processes, causal interactions, and conjunctive and interactive forks. It remains to see how these fare on our minimal desiderata for any adequate theory of probabilistic causation. First, let us explain how (ASYMMETRY) is satisfied by conjunctive forks.

Causal processes and interactive forks are perfectly symmetric structures. In STPC, the asymmetry of the production of the causal structure of the world is only given by the structure of the conjunctive forks. Salmon indeed follows Reichenbach in claiming that conjunctive forks contain an essential asymmetry: they are always open toward the direction of causation. In other words, C is always a common cause and never a common effect. This amounts to saying that causal forks where two events produce a common effect in the absence of a common cause never satisfy relations that characterize conjunctive forks. Given a causal structure, the direction of causation is thus the direction toward which most conjunctive forks are open. Consider our examples: while heavy smoking fulfills the requirements for C in Definition 41, suffering from a heart attack, as a common effect, does not. So, the direction given by the structure of conjunctive fork is the one we were wanting to recover: from smoking to either yellow fingers or lung cancer, and from running and/or smoking to suffering from a heart attack.

It is clear that STPC satisfies (ASYMETRY). Note that the account of the causal order above does not appeal to the time order. It is not a priori that the causal and the temporal order match. Salmon leaves open the question whether there could be cases of backward causation. He takes it as a contingent fact that in our world, most of the conjunctive forks are open toward the future.[37]

[37]Note that Dowe takes it that Bell-type phenomena constitute instances of backward causation in [38, p.

Let us now turn to the issue of whether or not STPC satisfies (ISOLATION). First, spurious causes can be dismissed simply because there is no causal process between a spurious cause and its putative effect. Recall the spurious correlation example that involved heavy smoking, yellow fingers and lung cancer. We were worried that the yellow fingers could screen off lung cancer from heavy smoking, just as heavy smoking screens off lung cancer from yellow fingers so the event of having yellow fingers could count as a cause for lung cancer and heavy smoking. Clearly, the consideration of causal processes could do at least part of the work. There is a causal process between the possession of yellow fingers and heavy smoking: plausibly, some chemical substance produced by the combustion of the cigarette which turns fingers yellow. However, there is no such causal process between the yellow fingers and lung cancer. No conserved quantity is transmitted and exchanged between the fingers and the lungs. By contrast, there is one such exchange between a smoked cigarette and the lungs of a smoker – through the smoke, and at least the depositing of tar in the lungs. The fact that causal processes must exist between causes and effects is sufficient to bypass the problem of spurious causes.

Recall now the problem of cancellations of correlations: it could happen that the average smokers are found less likely to suffer from heart attack than the average non-smokers for smokers are also more likely to be regular runners. Considering causal processes allows us to discover that there are separate causal processes between smoking and suffering from heart attacks and between running regularly and suffering from heart attacks. However, there is no such causal process between smoking and running regularly.

In short then, STPC, equipped with the apparatus of causal processes and the structure of conjunctive forks, meet the two minimal desiderata for any adequate theory of probabilistic

186]. He does not provide an argument for this claim. It is an assumption. We shall not discuss his analysis. Instead, we shall try to apply Salmon's view on Bell-type situations. The reason for this is that, since we are exploring the various possible interpretations for the hidden causal picture underlying the Bell-type phenomena, it would be begging the question to assume right from the outset that such phenomena are instances of backward causation.

causation. Hence, they are good candidates, in addition to MPTC, for investigating the causal structure underlying Bell-type situations.

7.3 The respective hidden causal picture according to MPTC and STPTC

We have described how MPTC and STPC can be considered as good candidates for being adequate theories of probabilistic causation. With this in hand, we want to apply these two types of theories of probabilistic causation to Bell-type situations. More precisely, we want to assess whether or not (¬PI-CAUS)* – that a violation of PI* is indicative of a underlying causal relationship while a violation of OI* is not – and (¬OI-HOL)* – that a violation of OI* is indicative of a form of holism – hold in these theories. We shall maintain that (¬PI*-CAUS), and hence (¬OI*-HOL), hold in neither MPTC nor in STPC, at least in the strong sense that the "experimental metaphysics" program seems to favor. That said, within both theories, there are ways to maintain a weakened version of the mainstream interpretation.

7.3.1 The hidden causal picture according to MPTC

Recall from Subsection 7.2.2 that, according to interventionist MTPC, a variable X counts as a cause of Y, with respect to a given set of variables S, if and only if a change in the values of X obtained by an intervention results in a corresponding change in the values of Y.

Two points are worth making clear for our analysis. First, what MTPC give us is an empirical criterion for causation. As explained before, MTPC do not aim at a metaphysical account of causation. At best, MTPC remain agnostic as to the ontological status of physical relations corresponding to what we call causal relations. MTPC are only concerned with empirically observable regularities. A consequence of such an empiricist stance is that, whatever the result of the application of MTPC to Bell-type situations, MTPC do not issue metaphysical conclusions about causal links at the ontological level.

A second point to keep in mind is that, according to MPTC, a variable is always a cause of another variable *with respect to a set of variables.* So, the first step in giving a MTPC analysis of Bell-type phenomena is to specify the set of variables considered. In Subsections 4.3.1 and 4.3.2, we contrasted the traditional picture of the causal structure underlying such phenomena with a completed picture, in which one takes the hidden variables of the apparatuses into account. Arguably, a manipulability analysis of Bell-type phenomena requires that we consider such a completed picture. It is a necessary condition for using the apparatus of interventions that we have a full record of all the possible interferences with the correlations studied. Thus, our set of variables contains not only A, B, I, J, and Λ, but also Γ and Δ.[38]

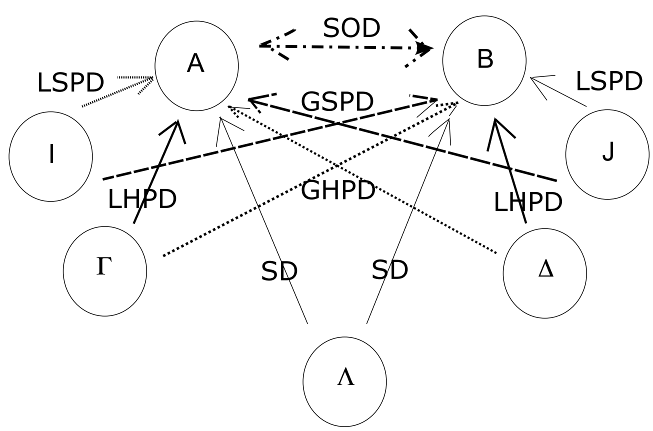

Figure 7.2: The various causal paths potentially responsible for the correlations in Bell-type phenomena.

[38]Consistent with our previous notation, the variable Γ and Δ respectively take $\{\gamma, \gamma', ...\}$ and $\{\delta, \delta', ...\}$ as values.

In Figure 7.2, we have represented all the possible causal paths from the set of variables we consider, $\mathcal{S}$. The arrows stand for potential direct causal influence (positive or negative). And we have labeled the causal paths as follows:

- SOD: Surface Outcome Dependence: dependence between outcomes, everything else held fixed – this will correspond to ¬OI*,
- LSPD: Local Surface Parameter Dependence: dependence between the outcome at one end and the surface parameter at the same end, everything else held fixed,
- GSPD: Global Surface Parameter Dependence: dependence between the outcome at one end and the surface parameter at the other end, everything else held fixed,
- LHPD: Local Hidden Parameter Dependence: dependence between the outcome at one end and the hidden parameter at the same end, everything else held fixed,
- GHPD: Global Hidden Parameter Dependence: dependence between the outcome at one end and the hidden parameter at the other end, everything else held fixed,
- SD: State Dependence: dependence between the outcomes and the state at the source.

Now that we have our set of variables and the putative causal relationships, we can apply the apparatus of interventions. Let us consider that we are looking for possible causes of the outcome variable B.[39] Now, it will be recalled that a variable X of $\mathcal{S}$ qualifies as a cause of B if and only if, were we to undertake an intervention on X, we would obtain invariant changes in B. Remember the constraints on interventions. First, the intervention on X changes X. Second, nothing else changes it. Third, the intervention on X cannot influence B but through X: there are no other routes between the intervention and the putative effect under consideration. With this in hand, we can assess whether or not failure of PI* and/or OI* are indicative of causal relations.

Let us start with PI*. PI* states that the probability of the outcome at one end does not change under changes upon the parameters at the other end of the experiment (hidden *and*

[39]Everything we say about B will apply to A as well because of the symmetries inherent in Bell-type phenomena.

surface). So, failure of PI* consists in that some changes in the probability of the outcome at one end occur upon changes of surface or hidden parameters at the other end. In the case under consideration, changes imposed on I or on Γ would result in changes in B. Now, this is supposed to indicate a causal path from the hidden or surface parameters at one end to the outcomes at the other end of the experiment. We maintain that this is not the case within a rigorous application of MTPC. Within MTPC, failure of PI* does not necessarily indicate a causal link between the parameters at one end and the outcomes at the other end of the experiment.

To see this, consider that while the first two requirements for interventions may be satisfied, the third is violated. Granted, assuming violation of PI*, one has the means to enact changes in B in changing I or Γ, while blocking interfering influences on B from λ, δ and j. The problem is that changes on I or on Γ can change B not only directly, but also through Local Parameter independence (hidden or surface), A and failure of OI*, or SOD. This point can be seen more easily in Figure 7.2. So, failure of PI* alone does not give us the means to assess whether there is a direct causal link from the parameters at one end to the outcomes at the other end. One has to make further assumptions about which causal paths in our causal picture are allowed or not. Only under the further assumption that either Local Parameter Dependence* or Outcome Dependence* is forbidden could one consider that failure of PI* is indicative of an underlying causal link between the parameters at one end of the experiment and the outcome at the other end.

Let us turn to OI* and its potential failure. Contrary to failure of PI*, failure of OI* seems to properly single out a unique causal path. Indeed, failure of OI* consists in that the effect B changes under changes on A, where every other potential factor is held fixed. In particular, conditionalizing on $\{i, \gamma, j, \delta\}$ forbids both influences on A from the parameters on the B side and influences on B from parameters on the A side. So, an intervention on A would be the only cause of change in A, and these changes imposed on A could not influence

B through any other route than the one linking the outcomes.

Here, however, arises a difficulty of a different sort, namely: is it possible to intervene on A at all? Of course, MTPC do not require that interventions can actually be done by us – that would be anthropocentric. That said, by definition, I counts as an intervention on A only if I causes A. Now, assuming that we have the complete set of all possible causal factors for A in our causal picture, and given that all these factors are maintained fixed in violations of OI*, then there is nothing left to be used as a trigger for an intervention on A. In short then, failure of OI* *alone* fails to give us the means to assess whether there is a causal link between the outcomes because no intervention seems to be possible on the outcome. Here again, one has to make further assumptions about which causal paths in our causal picture are allowed or not.

Three types of factors could be potentially used as triggers on A:

1. the parameters I and Γ on the A-side,

2. the parameters J and Δ on the B-side,

3. the state Λ.

Using 3. would obviously fail to satisfy the requirement of intervention. Indeed, if we assume that Λ causally influences A, we have arguably to assume that it influences B as well. In this case, our intervention can influence the putative effect through a route which does not pass through the putative cause, which prevents one from concluding that failure of OI* is indicative of a causal relationship.

What if we assume in our model that we can use 2. – i.e. the distant parameters, in order to trigger A? This amounts to assuming that Global Parameter Dependence, hidden or surface, holds.[40] In order to satisfy the requirements for interventions, one has to make the

[40]Strictly speaking, we would need only Global Parameter Dependence from the B-side to the A-side. This would seem rather ad hoc however. Further, consider that one can make the assumption that GPD

additional assumption that changes on J and Δ do not change directly B. In other words, one would have to forbid Local Parameter Dependence, both hidden and surface.[41] So, in a model such that Global Parameter Dependence holds but Local Parameter Dependence is forbidden, violation of OI* would be indicative of a causal path between the outcomes.

Let us turn to the third case, that is, the case in which we assume that we can intervene on A through I or Γ. This amounts to assuming Local Parameter Dependence, hidden or surface. Conditionalization on Λ, J and Δ allows one to satisfy the second requirement for intervention that nothing other than I or Γ, can influence A. Finally, in order to satisfy the third condition on interventions, we would have to forbid any other route from I or Γ to B. Thus, we have to forbid Global Parameter Dependence, both surface and hidden. So, in a model such that Local Parameter Dependence, either hidden or surface, holds, while Global Parameter Dependence, both hidden and surface, are forbidden, failure of OI* counts as indicative of an causal link between the outcomes.[42]

If we consider a model in which only OI* is violated, while any kind of Parameter Dependence, Local and Global, Hidden or Surface, is forbidden, then one has to admit that no intervention is possible on A. In turn, according to MTPC, in such a model, and relative to the set of variables considered, failure of OI* is not indicative of a causal link between the outcomes, or there is no causal link which has empirical consequences. Granted, it might well be the case that there is a causal link at the ontological level – if there are such things and whatever they may be – but this is not what MTPC are interested in. As said before, all MTPC care about are empirically observable regularities. From such a point of view, there is no causal relation when there is no empirical manifestation.

holds on both sides, since we already forbid that the parameters on the A-side influence B by holding fixed on i and γ.

[41] As in the previous case, we need strictly speaking to assume only that Local Parameter Dependence be forbidden on the B-side. The case in which it is forbidden on both sides can be considered however because, again, i and γ are held fixed in our model.

[42] Note that this option corresponds to Jones and Clifton's theorem – See Theorem 1 in Subsection 4.3.4.

Putting our analysis of violation of OI* and violation of PI* together, we can finally conclude that a rigorous application of MTPC supports a refined version of the mainstream interpretation: according to MTPC, in a model in which any kind of Parameter Dependence (Local and Global, Hidden or Surface) is forbidden, but only Outcome Independence is violated, there is a causal link between the parameters at one end of the experiment and the outcomes at the other end, whereas there is no causal link between the outcomes.

Note however that the version of the mainstream interpretation that MTPC can support is, however, weakened. It holds only from the empirical point of view: no metaphysical conclusions about the fundamental causal picture can be drawn from the MTPC. For the same reason, whether or not failure of OI* is best interpreted in terms of holism is left open within the empirical point of view of MTPC. That said, for defendants of the mainstream interpretation that are not too metaphysically minded, MTPC offers rigorous support.

7.3.2 The hidden causal picture according to STPC

It remains to be seen whether or not the traditional interpretation holds, or at least can be made plausible within STPC. We shall argue in this final subsection that under a strict application of STPC, this is not the case. That said, we shall also argue that, with some refinements of the theory, (¬PI-CAUS)* and (¬OI-HOL)* can be made plausible, but in a way which would here again not content Jarrett and Shimony's metaphysical aims.

Once again, recall that (¬PI-CAUS)* states that failure of PI* is indicative of an underlying causal relationship between the parameters at one end of the experiment and the outcomes at the other end, while failure of OI* is not indicative of causal relationship between the outcomes. In order to assess whether this holds or not within STPC, the question is then: how can we assess whether or not there is a causal relationship between two events with STPC?

It will be recalled from Subsection 7.2.3 that, according to STPC, the notion of causal relationship between events is not fundamental. Instead, the causal structure of the world is fundamentally constituted of:

1. causal processes;

2. causal interactions, including interactive and conjunctive forks.

Causal processes are responsible for the propagation of causal influences. Causal interactions, that is, interactive and conjunctive forks, are responsible for the modification of these propagating causal influences. Finally, causal interactions are defined in terms of intersection of causal processes.

We would like to make a special note from the outset to the effect that, if ¬PI* and ¬OI* are to be taken as indicative of underlying causal interactions, such interactions will have the structure of the conjunctive forks and not of the interactive forks. It will be recalled that, within the conjunctive fork structure, a common cause z satisfies the following condition:

$$P(x.y|z) = P(x|z).P(y|z), \tag{7.6}$$

which is a condition of independence, and which entails the screening off conditions:

$$P(x|y, z) = P(x|z) \quad \mathit{and} \quad P(y|x, z) = P(y|z). \tag{7.7}$$

The structure of interactive forks differs only in that, instead of 7.6, common causes within interactive forks satisfy 7.8:

$$P(x.y|z) > P(x|z).P(y|z). \tag{7.8}$$

It should be clear from the above that both PI* and OI* can be seen as requiring that the complete state λ screen off the parameters and the outcomes, or the two outcomes from one another, respectively. Hence, the traditional analysis clearly conceives Bell-type situations in terms of conjunctive, and not interactive, forks.

Now, according to the above, before we can decide whether or not there are causal interactions underlying ¬PI* and ¬OI*, we need to decide what counts as a causal process in Bell-type situations. Definition 39 in Subsection 7.2.3 states that a causal process is the worldline of an object which transmits a conserved quantity, the function of which, in the case of probabilistic causation, is to transmit determinate probability distributions over outcomes of possible interactions.

From this characterization, it follows that what counts as a potential causal process depends first on what counts as a worldline. To the best of our knowledge, Salmon is not clear upon how he interprets relativistic spacetimes that STPC rely on.[43] We shall then consider two options here:

(1) spacelike wordlines are not allowed;

(2) spacelike wordlines are allowed.

If spacelike wordlines are not allowed, then the only candidates for being causal processes are the two worldlines from the source to each of the apparatuses. In this case, the only possible causal pictures for Bell-type situations seem to be:[44]

(1a) the state λ is a common cause for the outcomes;

(1b) there is backward causation.

Arguably, Bell-type experiments rule out (1a): there is local common cause model for quantum probabilities. So, the upshot of Bell-type experimental results under STPC which do not allow for spacelike worldlines is that we have a case of backward causation. It will be recalled that this is consistent with Salmon's claim that the direction of causation is given by

[43]Salmon makes clear that continuity in spacetime is necessary, but this does not forbid superluminal processes.

[44]The question here arises as to whether all correlations need to be explained within STPC. Salmon gives explicitly a positive answer to this. Typically, he takes that correlations have to be explained in terms of a common cause, for example in [99, p.137]. One could of course consider a spacetime theory of probabilistic causation in which not all correlations have to be explained in terms of a common cause. That said, it is famously difficult to design a satisfactory criterion for distinguishing between correlations that have to be explained from correlations that do not. This relates to the conditions of applicability of the PCC, which we shall discuss later on in this chapter.

the direction of conjunctive forks, but that it is not necessarily pointing toward the future. STPC thus does not rule out the possibility of backward causation.

One could consider various models of backward causation for Bell-type phenomena. One such model considers that a first future-directed causal process deploys from the source to the apparatus where the first measurement-interaction occurs. This interaction results in the production of a past-directed causal process[45] toward the source with which the past-directed process interacts. This interaction in turn results in the production of a future-directed causal process from the source toward the second apparatus where the second measurement interaction occurs. We shall not investigate backward causation models in more detail. For our purposes, it is enough to conclude that, in the case in which spacelike worldline are not allowed, the rigorous application of STPC supports the interpretation of Bell-type situations in terms of backward causation. It cannot however support the mainstream interpretation.

Let us now consider Case (2), that is to say, the case in which spacelike worldlines are allowed in our model – say by ways of tachyons. In this case, not only could there be causal processes between the source and each of the apparatuses, but also between the parameters at one end of the experiment and the outcomes at the other end as well as between the outcomes. Given Jones and Clifton's theorem, we can, at least in principle, obtain transmissions of determinate probability distributions between I, Γ and b if PI* fails, as well as between a and b if OI* fails. So, under STPC, there can be some causal processes between outcomes or between outcomes and parameters, or both. Under STPC, both ¬PI* and ¬OI* could be indicative of causal interactions.

What are the various causal pictures that one could consider under STPC given these causal processes? For failure of OI*, it seems that the three following pictures are possible:

(2a) the state λ is a common cause of the outcomes;

(2b) there is direct (tachyonic) causation between the outcomes;

[45]For details about what such causal processes could be, see [38, p. 186 sq.] and reference therein.

(2c) there is backward causation.

Again, Bell-type experiments rule out (2a). So that the upshot of Bell-type experimental results under STPC which allow for spacelike wordline is that we have either tachyonic or backward causation. It should be clear that in none of these cases is the mainstream interpretation supported under STPC. In case there is backward causation, it is not supported because there are no causal links; neither between the parameters and the outcomes, nor between the outcomes. In case there are tachyonic causal processes, such processes can be between parameters and outcomes as well as between outcomes. In this case, failure of PI* as well as failure of OI* would both amount to superluminal causation, and to causal anomalies if superluminal causation is forbidden for other reasons.

The upshot of our analysis is that, under the strict application of STPC, (¬PI-CAUS)* fails, and hence (¬OI-HOL)* fails as well. Either failure of OI* and failure of PI* are indicative of underlying causal interactions, or none of them are. Thus, STPC does not support the mainstream interpretation.

It might be objected that STPC are at best incomplete, at worst unacceptable accounts of causality. The treatment of correlated events is but a slightly refined version of Reichenbach's Principle of Common Cause (PCC): any common cause as characterized by the conjunctive fork structure is a common cause as defined by the PCC. Now, the PCC is well known to fail in general.[46] More precisely, there are many types of correlations which fail to fulfill the requirements of the PCC. Moreover, such failures of the PCC are "benign"; they are not considered peculiar causal anomalies. In short then, not all correlations should be taken as having to satisfy the requirements of the PCC in any adequate account of causality. This does not mean that the PCC is not important or useful. It just means that the conditions of application of the PCC have to be refined. One thus could argue that, similarly, STPC must be refined such that the domain of application of the conjunctive fork structure be made

[46]For a more detailed presentation of the PCC, see Subsection 6.4.1.

more restricted.

To elaborate in detail the various types of counterexamples that have been leveled against the PCC in the literature would take us too far off track.[47] We shall not attempt to provide conditions of applicability for the PCC in general. We shall instead consider one type of failure of the PCC which is of interest for the analysis of Bell-type situations.

Consider a flock of birds flying. Each bird is separated in space from the others. It is well known that there are some species for which there is no leader of the flock. However, the trajectories of the individual birds are highly correlated. There is no common cause that would screen off such correlations. Hence, the PCC fails for such correlations. In this, as in similar cases, the diagnosis for the problem with the application of the PCC is that it is inappropriate to explain correlations in events when they can be seen as being events that characterize a single system. As Frank Arntzenius writes:[48]

> Thus, [...] we regard such systems as single unified systems, and do not demand a common cause explanation for the correlated motions and properties of their parts. A fairly intuitive notion of what counts as a single system, after all, is a system that behaves in a unified manner, i.e. a system whose parts have a very strong correlation in their motions and of other properties, no matter how complicated the set of influences acting on them. For instance a rigid physical object has parts whose motions and properties are strongly correlated, no matter how complicated the influences acting on it. These systems therefore are naturally and usefully treated as single systems for almost any purpose.

If that is right, then an argument could be made to support the traditional interpretation of Bell-type situations. One can argue that correlations due to failure of OI* correspond

[47]The literature here is plentiful. See, among others, [128], [25], [105, 106], [126], [90], [42, chap. 5], [127], and [19]. See also the work done by Szabo and his team, for example [67, 68]. Finally, Frank Arntzenius has a series of articles ([2, 3, 4]) and is the author of the recent synthesis on the principle of common cause ([5]). We shall rely mostly on Frank Arntzenius' work in the following.

[48][5].

to such failures of the PCC. For the events a and b are measurement outcomes meant to reflect properties of two systems which originate from the same source. Correlations between the measurement results can been seen as resulting from interactions with a single system. In that case there would be no need to conclude that there is a causal influence between outcomes or a case of backward causation.

Now, considering failure of PI*, no such argument could be made. To regard the parameters and the outcomes as properties of a single system seems both unnatural and problematic. It seems unnatural because we have good reason to believe that the apparatus and the quantum system that is measured are separated physical systems – we can use an apparatus on arbitrary quantum systems for example. It seems problematic because it would amount to denial of the very possibility of a scientific experiment in Bell-type situation by denying the distinction between the apparatus and the object of the experiment.

Notice, however, that it is unlikely that all defenders of the traditional interpretation would ultimately accept such an argument in support of their theses. The approach that we have advocated for the restriction of the application of PCC does not entitle us to any definitive ontological conclusions about the fundamental causal picture underlying Bell-type situations. First, what we have are only plausibility arguments to the effect that correlations by failure of OI* could be due to a form of holism. Second, the form of holism these arguments can support has strong epistemological leanings. Arntzenius' suggestions describe ways in which we can most successfully carve out the worlds into individual systems. The rule above tells us that certain correlated events *can be regarded* as properties of a single unified system, not that these correlated events *are* the properties of a single unified system at the ontological level. The flock of birds example makes the point explicit. The argument given here in favor of (¬PI*-CAUS) and (¬OI*-HOL) thus drastically weakens the traditional interpretation: it does not support the idea of "experimental metaphysics".

7.4 Conclusion

In this chapter, we have considered two theories of probabilistic causation, namely the interventionist manipulability theories and the spacetime theories of probabilistic causation (MTPC and STPC) in order to assess two components of the mainstream interpretation of Bell-type theorems and Bell-type phenomena, namely (¬PI-CAUS)* and (¬OI-HOL)*.

We have shown that under a strict application of MTPC, a good argument can be made in favor of (¬PI-CAUS)*. That said, MTPC deal strictly with causation at the empirical level and do not license inferences concerning causation at the ontological level. The mainstream interpretation is supported by MTPC only at the empirical level. Failure of PI* is indicative of a causal link with empirical consequences, but failure of OI* is not indicative of any causal link which can manifests itself as the empirical level. So MTPC support the idea of"peaceful coexistence" between models of quantum phenomena in which only OI* is violated while PI* holds and Relativity (if Relativity is taken to forbid superluminal causation) at the empirical level, but leave open what the world is like behind the phenomena. This is far from supporting any kind of "experimental metaphysics" but still could content many physicists and philosophers.

STPC provide a more metaphysically oriented account of causation. However, we have shown that under a strict application of STPC, the mainstream interpretation fails. Either failure of PI* and failure of OI* both count as indicative of underlying causal interactions, or neither of them do. It follows from this that no account of causation supports the mainstream interpretation at the ontological level.

However, we have argued that under a refined version of the spacetime theories of probabilistic causation, some plausibility arguments can be made in favor of a weakened version of the traditional interpretation, according to which failure of PI* is to be regarded as indicative of a causal interaction between the parameters and the outcomes, while failure of

OI* indicates that we should regard the two subsystems as parts of a single unified system.

The upshot is thus that neither of types of theories of probabilistic causation available can support the mainstream interpretation construed as a strong program of experimental metaphysics. That said, (¬PI-CAUS)*, if restricted to the empirical level, can be well supported within MTPC. Finally, (¬OI-HOL)*, if restricted to the epistemological level, is plausible within a refined version of STPC.

Bibliography and Tables

List of Definitions and Theorems

Cause: counterfactual account	Definition 11	Subsection 4.3.1
Cause: conditional probabilistic account	Definition 12	Subsection 4.3.1
Causal relevance: conditional probabilistic account	Definition 13	Subsection 4.3.1
Cause: genuine	Definition 34	Subsection 7.1.2
Cause: spurious	Definition 35	Subsection 7.1.2
Cause: manipulability account	Definition 36	Subsection 7.2.2
Intervention	Definition 37	Subsection 7.2.2
Causal interaction: spacetime theories	Definition 38	Subsection
Causal process: spacetime theories	Definition 39	Subsection 7.2.3
Transmission: spacetime theories	Definition 40	Subsection 7.2.3
Conjunctive forks	Definition 41	Subsection 7.2.3

List of Figures

Bibliography

[1] H. Araki and M. Yanase. Measurement of quantum mechanical operators. *Physical Review*, 120(2):622–26, 1960.

[2] F. Arntzenius. Physics and common causes. *Synthese*, 82:7796, 1990.

[3] F. Arntzenius. The common cause principle. *Philosophy of Science*, In Hull, D., M. Forbes, and K. Okruhlik, eds. PSA 1992, vol. 2. East Lansing: Philosophy of Science Association:22737, 1993.

[4] F. Arntzenius. Transition chances and causation. *Pacific Philosophical Quarterly*, 78/2:149–68, 1997.

[5] F. Arntzenius. Reichenbachś common cause principle. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2005.

[6] A. Aspect. *Trois tests expérimentaux des inégalités de Bell par mesure de corrélation de polarisation de photons*. Thèse d'Etat, Orsay, 1983.

[7] A. Aspect. Bell's theorems: the naive view of an experimentalist. In R. A. Bertlmann and A. Zeilinger, editors, *Quantum [UN]speakables – From Bell to Quantum Information*. Springer, Berlin Heidelberg New York, 2002.

[8] A. Aspect, J. Dalibard, and G. Roger. Experimental test of Bell's inequalities using time-varying analysers. *Physical Review Letters*, 49:1804–07, 1982.

[9] G. Bacciagaluppi. The role of decoherence in quantum mechanics. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2005 edition, 2005.

[10] J. Bell. Against 'measurement'. In *Speakable and Unspeakable in Quantum Mechanics*, pages 213–231. Cambridge University Press, United Kingdom, second edition, 2004.

[11] J. S. Bell. Introduction to the hidden-variable question. In *Speakable and Unspeakable in Quantum Mechanics*, pages 29–39. Cambridge University Press, Cambridge, 1987.

[12] J. S. Bell. On the Einstein-Podolski-Rosen paradox. In *Speakable and Unspeakable in Quantum Mechanics*, pages 14–21. Cambridge University Press, Cambridge, 1987.

[13] J. S. Bell. On the problem of hidden variables in quantum mechanics. In *Speakable and Unspeakable in Quantum Mechanics*, pages 1–13. Cambridge University Press, Cambridge, 1987.

[14] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, Cambridge, 1987.

[15] J. S. Bell. The theory of local beables. In *Speakable and Unspeakable in Quantum Mechanics*, pages 52–62. Cambridge University Press, Cambridge, 1987.

[16] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, Cambridge, (new edition) 2004.

[17] J. Berkovitz. Aspects of quantum non-locality I: Superluminal signaling, action-at-a-distance, non-separability and holism. *Studies in the History and Philosophy of Modern Physics*, 29(2):183–222, 1998.

[18] J. Berkovitz. Aspects of quantum non-locality II: Superluminal causation and relativity. *Studies in the History and Philosophy of Modern Physics*, 29(4):509–45, 1998.

[19] J. Berkovitz. The many principles of the common cause. *Reports on Philosophy*, 20:51–83, 2000.

[20] K. Brading and E. Landry. Scientific structuralism: Presentation and representation. *Philosophy of Science*, 73:571–581, 2006.

[21] J. Butterfield. A space-time approach to the Bell inequality. In James T. Cushing and Ernan McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 114–44. University of Notre Dame Press, Notre Dame, 1989.

[22] J. Butterfield. Bell's theorem: What it takes. *The British Journal for the Philosophy of Science*, 43(1):41–83, 1992.

[23] J. Butterfield and T. Placek, editors. *Non-Locality and Modality*. NATO science series: II: Mathematics, Physics. Springer, Berlin Heidelberg New York, 2002.

[24] R. Carnap. *Philosophical Foundations of Physics*. Macmillian, New York, 1966.

[25] N. Cartwright. Causal laws and effective strategies. *Noûs*, 13:419–37, 1983.

[26] N. Cartwright. *How the Laws of Physics Lie*. Oxford University Press, Oxford, 1983.

[27] N. Cartwright. *Natures Capacities and their Measurement*. Oxford University Press, Oxford, 1989.

[28] J. F. Clauser and M. A. Horne. Experimental consequences of objective local theories. *Physical Review D*, 10:526–35, 1974.

[29] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review D*, 10:52635, 1974.

[30] R. Clifton and C. Pagonis. Unremarkable contextualism: Dispositions in the Bohm theory. *Foundations of Physics*, 25(2):281–96, 1995.

[31] J. T. Cushing and E. McMullin, editors. *Philosophical Consequences of Quantum Theory*. University of Notre Dame Press, Notre Dame, 1989.

[32] N. da Costa and S. French. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford University Press, Oxford, 2003.

[33] H. W. de Regt and D. Dieks. A contextual approach to scientific understanding. *Synthese*, pages 137–170, 2005.

[34] A. de Saint Exupéry. *Le Petit Prince*. Reynald and Hitchcock, New York, 1943.

[35] W. M. Dickson. *Quantum Chance And Non-locality: Probability And Non-locality In The Interpretations Of Quantum Mechanics*. Cambridge University Press, Cambridge, 1998.

[36] P. Dowe. Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59:195–216, 1992.

[37] P. Dowe. Causality and conserved quantities: a reply to Salmon. *Philosophy of Science*, 62:321–333, 1995.

[38] P. Dowe. *Physical Causation*. Cambridge University Press, Cambridge and New York, 2000.

[39] P. Dowe. Chance-lowering causes. In P. Noordhof P. Dowe, editor, *Cause and Chance: Causation in an Indeterministic World*, pages 39–57. Routledge, London and New York, 2004.

[40] D. Dürr, S. Goldstein, and N. Zanghi. Quantum equilibrium and the origin of absolute uncertainty. *Journal of Statistical Physics*, 67, No 5-6, 1992.

[41] J. Earman. Locality, non-locality and action at a distance: A skeptical review of some philosophical dogmas. In P. Achinstein and R. Kargon, editors, *Theoretical Physics in the 100 Years since Kelvin's Baltimore Lectures*. MIT Press, Cambridge, 1986.

[42] J. Earman. *Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford University Press, Oxford and New York, 1995.

[43] H. Reichenbach (edited by M. Reichenbach). *The Direction of Time*. University of California Press, Berkeley, Los Angeles, 1956.

[44] H. Reichenbach (edited by M. Reichenbach). *The Philosophy of Space and Time*. Dover, New York, 1958.

[45] A. Fine. Logic, probability,and quantum theory. *Philosophy of Science*, 35:101–11, 1968.

[46] A. Fine. Probability in quantum theories and in other statistical theories. In Mario Bunge, editor, *Problems in the Foundations of Physics*, pages 79–92. Springer, Berlin Heidelberg New York, 1971.

[47] A. Fine. Probabilities and the interpretation of quantum mechanics. *British Journal for the Philosophy of Science*, 24:1–37, 1973.

[48] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physics Review Letters*, 48:291–5, 1982.

[49] A. Fine. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23:1306–10, 1982.

[50] A. Fine. Some local models for correlation experiment. *Synthese*, 50:279–94, 1982.

[51] A. Fine. *The Shaky Game, Einstein Realism and the Quantum Theory*. The University Chicago Press, Chicago, 1986.

[52] A. Fine. Do correlations need to be explained? In J. T. Cushing and E. McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 175–94. University of Notre Dame Press, Notre Dame, 1989.

[53] A. Fine. Locality and the Hardy theorem. In J. Butterfield and C. Pagonis, editors, *From physics to philosophy*. Cambridge University Press, Cambridge, 1999.

[54] B. Fogel. The structure of holism: what it takes for probability distributions to be separable and to factorize. *manuscript*, 2005.

[55] L. T. F. Gamut. *Logic, Language and Meaning: Introduction to Logic*. University of Chicago Press, Chicago, 1990.

[56] R. N. Giere. *Understanding Scientific Reasoning*. Holt, Rinehart and Winston, New York, 1979.

[57] A. M. Gleason. Measures on closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics*, 6:885–93, 1957.

[58] I. J. Good. A causal calculus i. *British Journal for the Philosophy of Science*, 11:305–308, 1961.

[59] I. J. Good. A causal calculus ii. *British Journal for the Philosophy of Science*, 12:43–51, 1961.

[60] R. A. Healey. *The Philosophy of Quantum Mechanics: an Interactive Interpretation*. Cambridge University Press, Cambridge, 1989.

[61] R. A. Healey. Holism and nonseparability. *Journal of Philosophy*, 88:393–421, 1991.

[62] R. A. Healey. Nonseparability and causal explanation. *Studies in History and Philosophy of Modern Physics*, 25:337–74, 1994.

[63] R. A. Healey. Holism and nonseparability in physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2004.

[64] G. Hellman. Stochastic Einstein-locality and the Bell theorems. *Synthese*, 53:461–504, 1982.

[65] C. G. Hempel. Implications of Carnap's work for the philosophy of science. In P. Schlipp, editor, *The Philosophy of Rudolph Carnap*, pages 685–709. Open Court Publishing, Lasalle, III., 1962.

[66] C. G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. Free Press, New York, 1965.

[67] G. Hofer-Szabó, M. Rédei, and L. E. Szabó. On Reichanbach's common cause principle and on Reichenbah's notion of common cause. *British Journal for the Philosophy of Science*, 5:377–99, 1999.

[68] G. Hofer-Szabó, M. Rédei, and L. E. Szabó. Common causes are not common common causes. *Philosophy of Science*, 69:623–33, 2002.

[69] D. Howard. Einstein on locality and separability. *Studies in History and Philosophy of Science*, 16:171–201, 1985.

[70] D. Howard. Holism, separability and the metaphysical implications of the Bell experiments. In J. T. Cushing and E. McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 224–53. University of Notre Dame Press, Notre Dame, 1989.

[71] R.I.G. Hughes. *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press, Cambridge, MA and London, England., 1989.

[72] J. P. Jarrett. On the physical significance of the locality conditions in the Bell arguments. *Noûs*, 18:569–89, 1984.

[73] J. P. Jarrett. Bell's theorem: A guide to the implications. In J. T. Cushing and E. McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 61–79. University of Notre Dame Press, Notre Dame, 1989.

[74] M. R. Jones and R. K. Clifton. Against experimental metaphysics. *Midwest Studies in Philosophy*, XVIII:295–316, 1993.

[75] S. Kochen and E. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87, 1967.

[76] H. Krips. Measurement in quantum theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2007.

[77] F. Kronz. Hidden locality, conspiracy and superluminal signals. *Philosophy of Science*, 57:420–44, 1990.

[78] J. A. Larsson and R. D. Gill. Bell's inequality and the coincidence-time loophole. *Europhysics letters*, 67:707–13, 2004.

[79] S. Le Bihan. La conception sémantique des théories scientifiques, en quoi a-t-elle pu se dire 'sémantique'? *Matière Première*, 1:215–250, 2006.

[80] D. Lewis. Causation. *Journal of Philosophy*, 70:556–567, 1973.

[81] D. Lewis. Causation as influence. *Journal of Philosophy*, 97:182197, 1973.

[82] L. Löwenheim. Über Möglichkeiten im Relativkalkül. In J. van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, pages 447–470. Harvard University Press, Cambridge MA, 1915 [1967].

[83] T. Maudlin. *Quantum Non-Locality and Relativity, Metaphysical Intimations of Modern Physics*. Blackwell, Oxford, 1994.

[84] P. Menzies and H. Price. Causation as a secondary quality. *British Journal for the Philosophy of Science*, 44:187203, 1993.

[85] N. D. Mermin. Generalizations of Bell's theorem to higher spins and higher correlations. In Roth L.M. and Inomato A., editors, *Foundamental Questions in Quantum Mechanics*, pages 7–20. Gordon and Breach, 1986.

[86] M. S. Morgan and M. Morrison. *Models as Mediators*. Cambridge University Press, Cambridge, 1999.

[87] M. Morrison. Models as autonomous agents. In M. Morrison M. S. Morgan, editor, *Models as Mediators*, pages 38–65. Cambridge University Press, Cambridge, 1999.

[88] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, 2000.

[89] H. Price. Agency and probabilistic causality. *British Journal for the Philosophy of Science*, 42:157176, 1991.

[90] H. Price, editor. *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press, Oxford and New York, 1996.

[91] H. Putnam. What theories are not. In A. Tarski E. Nagel, P. Suppes, editor, *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, pages 240–252. Stanford University Press, Stanford, 1962.

[92] M. Redhead. Models in physics. *British Journal for Philosophy of Science*, 31(1), 1980.

[93] M. L. G. Redhead. *Incompleteness, Nonlocality, and Realism, A Prolegomenon to the Philosophy of Quantum Mechanics.* Clarendon, 1987.

[94] M. L. G. Redhead. Nonfactorizability, stochastic causality and passion at a distance. In J. T. Cushing and E. McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 145–52. University of Notre Dame Press, Notre Dame, 1989.

[95] H. Reichenbach. The causal structure of the world and the difference between past and future. In M. Reichenbach and R. S. Cohen, editors, *Selected Writings 1909-1953*, volume 2, pages 81–117. Reidel, 1978.

[96] W. C. Salmon. Probabilistic causality. *Pacific Philosophical Quarterly*, 61:50– 74, 1980.

[97] W. C. Salmon. *Scientific Explanation and The Causal Structure of the World.* Princeton University Press, Princeton, 1984.

[98] W. C. Salmon. Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64:461–77, 1997.

[99] W. C. Salmon. *Causality and Explanation.* Oxford University Press, Oxford, 1998.

[100] A. Shimony. Metaphysical problems in the foundation of quantum mechanics. *International Philosophical Quaterly*, 18:3–17, 1978.

[101] A. Shimony. Contextual hidden variables theories and the Bell's inequalities. *British Journal for the Philosophy of Science*, 35(1):25–45, 1984.

[102] A. Shimony. Events and processes in the quantum world. In R. Penrose and C. Isham, editors, *Quantum Concepts in Space and Time*, pages 182–203. Oxford University Press, Oxford, 1986.

[103] A. Shimony. Search for a worldview which can accomodate our knowledge of microphysics. In J. T. Cushing and E. McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 25–37. University of Notre Dame Press, Notre Dame, 1989.

[104] A. Shimony. Bell's theorem. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2005/entries/physics-Rpcc/, Summer 2005.

[105] E. Sober. The principle of the common cause. In J. Fetzer, editor, *Probability and Causality*, pages 211–29. Dordrecht: Reidel, 1988.

[106] E. Sober. Venetian sea levels, british bread prices, and the principle of common cause. *The British Journal for the Philosophy of Science*, 52:331–46, 2001.

[107] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediciton and Search.* The MIT Press, Cambridge, MA, 1993, Sd. ed. 2000.

[108] M. Suarez. Scientific representation: Against similarity and isomorphism. *International Studies in Philosophy of Science*, 17(1):225–244, 2003.

[109] F. Suppe. The search for philosophical understanding of scientific theories. In F. Suppe, editor, *The Structure of Scientific Theories*, pages 3–232. University of Illinois Press, Urbana, 1974.

[110] F. Suppe. *The Semantic Conception of Theories and Scientific Realism.* University of Illinois Press, Urbana and Chicago, 1989.

[111] P. Suppes. Models of data. In P. Suppes, editor, *Studies in the Methology and Foundations of Science, Selected Papers from 1951 to 1969*, pages 24–35. Reidel, 1962.

[112] P. Suppes. What is a scientific theory? In S. Morgenbeser, editor, *Philosophy of Science Today*, pages 55–67. Basic Books, New York, 1967.

[113] P. Suppes. A comparison of the meaning and uses of models in mathematics and the empirical sciences. In *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, pages 10–23. Reidel, Dordrecht, 1969.

[114] P. Suppes. *A Probabilistic Theory of Causality.* North Holland, Amsterdam, 1970.

[115] P. Suppes. Answer to Joseph Sneed on structural explanantion. In P. Humphreys, editor, *Patrick Suppes : Scientific Philosopher, Vol 2, Philosophy of Physics, Theory Structure, and Measurement Theory*, volume 234, pages 213–16. Kluwer, Synthese Library, 1994.

[116] P. Suppes. *Representation and Invariance of Scientific Structures, CSLI Lecture Notes.* Center for the Study of Language and Information, Stanford, 2002.

[117] P. Suppes and M. Zanotti. On the determinism of hidden variable theories with strict correlations and conditional independence of observables. In P. Suppes and M. Zanotti, editors, *Foundations of Probability with Applications*, pages 83–91. Reidel, 1976.

[118] G. Svetlichny, M. Redhead, H. Brown, and J. Butterfield. Do the Bell inequalities require the existence of joint probability distributions? *Philosophy of Science*, 55:387–401, 1988.

[119] L. E. Szabo. On Fine's resolution of the EPR-Bell's problem. *Foundations of Physics*, 30:1891–1909, 2000.

[120] A. Tarski. The concept of truth in the languages of the deductive sciences. In J. Corcoran, editor, *Logic, Semantics, Metamathematics, papers from 1923 to 1938*. Hackett Publishing Company., Indianapolis, 1933.

[121] P. Teller. Quantum mechanics and the nature of continuous physical quantities. *Journal of Philosophy*, 76,7:345–62, 1979.

[122] P. Teller. The projection postulate of quantum mechanics: A new perspective. *Philosophy of Science*, 51,3:369–95, 1984.

[123] P. Teller. Relational holism and quantum mechanics. *British Journal for the Philosophy of Science*, 37:71–81, 1986.

[124] P. Teller. Relational holism and the Bell's inequality. In James T. Cushing and Ernan McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 208–18. University of Notre Dame Press, Notre Dame, 1989.

[125] M. Thomson-Jones. Models and the semantic view. *Philosophy of science*, 73:524–535, 2006.

[126] R. Torretti. Do conjunctive forks always point to a common cause? *The British Journal for the Philosophy of Science*, 38:384–387, 1987.

[127] J. Uffink. The principle of the common cause faces the bernstein paradox. *Philosophy of Science*, 66, Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part I: Contributed Papers:S512–S525, 1999.

[128] B. C. van Fraassen. *The Scientific Image*. Oxford University Press, 1980.

[129] B. C. van Fraassen. The charybdis of realism. *Synthese*, 52:25–38, 1982.

[130] B. C. van Fraassen. The charybdis of realism: epistemological implications of Bell's inequality. In James T. Cushing and Ernan McMullin, editors, *Philosophical Consequences of Quantum Theory*, pages 97–113. University of Notre Dame Press, Notre Dame, 1989.

[131] B. C. van Fraassen. *Quantum Mechanics: An Empirist View*. Clarendon Press, Oxford, 1991.

[132] J. von Neumann. *Mathematical Foundations of Quantum Mechanics.* Princeton University Press, Princeton, 1955.

[133] G. von Wright. *Explanation and Understanding.* Cornell University Press, Ithaka, NY, 1971.

[134] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger. Violation of bell's inequality under strict einstein locality conditions. *Phys. Rev. Lett.*, 81:5039–43, 1998.

[135] E. Winsberg and A. Fine. Quantum life: interaction, entanglement, and separation. *Journal of Philosophy*, C(2):80–97, 2003.

[136] J. Woodward. *Making things happen : a theory of causal explanation.* Oxford University Press, Oxford, 2003.

[137] J. Woodward and C. Hitchcock. Explanatory generalization, part I: A counterfactual account. *Noûs*, 37(1):1–24, 2003.

[138] J. Woodward and C. Hitchcock. Explanatory generalization, part II: Plumbing explanatory depth. *Noûs*, 37(2):181–199, 2003.

[139] A. W'́uthrich. *Quantum Correlations and Common Causes.* Bern Studies in the History and Philosophy of Science, Bern, 2004.

[140] H. D. Zeh. Basic concepts and their interpretation. In *Decoherence adn the appearance of a classical world in quantum theory*, pages 7–40. Springer-Verlag, Germany, 2003.

[141] A. Zeilinger. Experiment and the foundations of quantum physics. *Review of Modern Physics*, 71:S288–97, 1999.